

Universidad Carlos III de Madrid

Escuela Politécnica Superior

Ingeniería Técnica de Telecomunicaciones:
Sistemas de Telecomunicación.



Selección causal de características para la
discriminación de arritmias cardiacas.

Proyectando: Francisco de Borja de Riva Solla.

Tutor: Mario de Prado Cumplido.

1 de Octubre de 2009.

Proyecto de Fin de Carrera

Selección causal de características para la discriminación de arritmias cardiacas.

La defensa del presente proyecto ha sido realizada el día 1 de Octubre de 2009, siendo calificada por el siguiente tribunal:

PRESIDENTE: MATILDE PILAR SÁNCHEZ FERNÁNDEZ

SECRETARIO: JERÓNIMO ARENAS GARCÍA

VOCAL: JOSE JESÚS GARCÍA RUEDA

habiendo obtenido la siguiente calificación:

CALIFICACIÓN:

'No os avergoncéis por caer en el camino. Toda persona ha caído alguna vez. Todos saben lo que es ser derrotado. Caerse y no levantarse sería algo por lo que estar avergonzado. Los mejores lo darán todo, y se volverán a levantar. Los corrientes tardarán en levantarse. Mientras que los perdedores se quedarán tumbados en el suelo.'

Darrel Royal, Texas, 1960.

'Si nuestra vida es tan breve como unos fuegos artificiales, hagamos que nuestro momento de brillo supere el de cualquier estrella.'

Agradecimientos:

Muchas personas han hecho posible este momento. Algunos siguen aquí, mientras que otros se han ido ya, pero no por ello son menos importantes:

En primer lugar darle las gracias a mi familia: mis padres y mis hermanos. Por haber estado ahí siempre, ayudándome en los momentos más duros, alegrándose con mis éxitos y apoyándome cuando lo he necesitado.

A mis amigos íntimos: David, Dani, Rocío, Marta, Margot, Mark ... Muchos habeis estado conmigo a cada paso del camino, siempre escuchando y dándome muy buenos consejos, os lo agradeceré siempre.

A mis compañeros de clase: Marta, Natalia, Miguel, Alberto... Tantas horas compartidas con prácticas, trabajos, memorias, resolviendo dudas de última hora. Han sido años difíciles, pero gracias a vosotros han sido siempre divertidos, entretenidos y sobre todo, inolvidables.

A mis amigos y maestros de Aikido: Muchas horas hemos estado juntos entrenando y tantas cosas hemos compartido que estarán ahí siempre.

A mis profesores, a todos ellos y en particular a Jerónimo, Mati y Jesús: por todas las dudas que me solucionasteis en estos años, por esa ayuda a la hora de decidir que hacer y por tantas preguntas que resolvisteis siempre sin perder la paciencia.

A mi tutor: Mario. Por ayudarme en los momentos en los que estaba atascado, siempre capaz de hacerme un hueco sin importar las fechas que fueran para sacar el proyecto adelante.

A todos aquellos que me habeis ayudado a estar aquí hoy, sin excepción alguna,

Gracias a todos.

Índice

I	Objetivos	17
1.	Fisiología	19
1.1.	Estructura básica del corazón:	20
1.1.1.	Estructura del corazón:	20
1.1.2.	Cavidades cardiacas.	21
1.1.3.	Ciclo Cardíaco:	22
1.2.	Trastornos en la velocidad del ritmo cardíaco	24
1.2.1.	Taquicardia sinusal	25
1.2.2.	Taquicardia supraventricular	26
1.2.3.	Taquicardia ventricular	27
2.	Selección causal de características	29
2.1.	Objetivo de la seleccion causal	29
2.2.	Fortalezas y debilidades	29
2.3.	Posibles combinaciones	30
2.4.	Descripción	30
2.5.	Un prototipo de algoritmo de descubrimiento causal	34
3.	Clasificación mediante decisión máquina	37

II	Herramientas	41
4.	Árboles Causales y Manto de Markov (Markov Blanket)	42
4.1.	Manto de Markov	42
4.2.	HITON MB e IAMB	44
5.	SVM: Support Vector Machine	47
5.1.	Forma primal	50
5.2.	Forma dual.	51
5.3.	Hiperplanos sesgados e insesgados	51
5.4.	Extensiones a la SVM lineal	52
5.5.	SVM Multiclase	53
5.5.1.	Implementación	54
6.	LIBSVM	55
6.1.	Selección de parámetros	61
III	Problema Médico	63
7.	Descripción	63
8.	Resultados del estudio	65
8.1.	Análisis de discriminación	66
8.1.1.	Capacidad de discriminación de la FPPI-V.	66
8.1.2.	Capacidad de discriminación de parámetros relacionados con el FPPI.	66

IV	Plan de Pruebas	69
9.	Filtrado de datos válidos	69
10.	Configuraciones	70
11.	Aplicación de algoritmos de selección de características	72
11.1.	Relación causal de variables	74
V	Resultados experimentales del proyecto.	77
12.	Réplica del experimento de clasificación	77
12.1.	Réplica usando SVM	77
12.1.1.	Condiciones suficientes	77
12.1.2.	Todas las variables disponibles	78
12.2.	Evaluación del conjunto de condiciones suficientes	78
12.3.	Evaluación del conjunto completo de variables	78
13.	Resultados experimentales de la selección de características	80
13.1.	Aplicando el algoritmo IAMB al conjunto previo	80
13.2.	Aplicando el algoritmo HITON_MB al conjunto previo	81
13.3.	Aplicando el algoritmo IAMB al conjunto de completo de variables	81
13.4.	Aplicando el algoritmo HITON_MB al conjunto completo de variables	81

14.Resultados de la obtención del árbol causal	83
14.1. Algoritmo TPDA aplicado al conjunto completo de variables	84
14.2. Algoritmo PC aplicado al conjunto completo de variables.	85
14.3. Algoritmo TPDA aplicado al conjunto previo	86
14.4. Algoritmo PC aplicado al conjunto previo	87
 VI Conclusiones	 89
 VII Líneas futuras de investigación	 93
 VIII Bibliografía	 95

Índice de figuras

1.	Estructura cardiaca del corazón	22
2.	Ejemplo de taquicardia sinusal	26
3.	Ejemplo de Taquicardia supraventricular	27
4.	Ejemplo de taquicardia ventricular	28
5.	Manto de Markov aplicado a la variable Lung Cancer	43
6.	Ejemplo de fronteras de clasificación.	48
7.	Ejemplo de dos hiperplanos paralelos al hiperplano frontera de margen máximo.	49
8.	Situación inicial del ejemplo SVM.	57
9.	Frontera de decisión óptima en caso simétrico	57
10.	Frontera de decisión ponderada para la clase 1	58
11.	Frontera de decisión ponderada para la clase 0.	59
12.	Frontera de decisión con factor de extensión de gaussianas y coste simétrico.	59
13.	Frontera de decisión con coste, ponderación y extensión de gaussianas.	60
14.	Frontera de decisión sobreajustada	61
15.	Gráfica representando la probabilidad de detección y de falsa alarma del algoritmo predictivo	67
16.	Resumen de la selección de características.	80

17.	Árbol causal para el conjunto completo de variables usando el algoritmo TPDA.	84
18.	Árbol causal para el conjunto completo de variables usando el algoritmo PC.	85
19.	Árbol causal para las condiciones suficientes usando el algoritmo TPDA. . .	86
20.	Árbol causal para las condiciones suficientes usando el algoritmo PC. . . .	87

Índice de cuadros

1.	Tabla de Parámetros configurados para las clasificaciones SVM en la evaluación de resultados.	71
2.	Tabla con los vectores de opciones para las clasificaciones.	71
3.	Tabla de parámetros para los algoritmos de selección de características, conjunto previo.	73
4.	Tabla de parámetros para los algoritmos de selección de características, conjunto completo.	73
5.	Expresiones para el cálculo de MB para el conjunto previo.	74
6.	Expresiones para el cálculo de MB para el conjunto completo de variables.	74
7.	Tabla de asignación de valores para la relación causal de variables aplicado al conjunto previo	75
8.	Expresiones de los árboles causales para el conjunto previo.	75
9.	Tabla de asignación de valores para la relación causal de variables aplicado al conjunto completo.	76
10.	Expresiones de los árboles causales para el conjunto completo.	76
11.	Resultados obtenidos con el conjunto previo.	77
12.	Resultados obtenidos a través del conjunto completo de variables.	78
13.	Resultados experimentales según el tipo de clasificador el conjunto previo.	78

14.	Resultados experimentales según el tipo de clasificador el conjunto completo.	79
16.	Resultados de la validación de la selección de características IAMB.	80
17.	Resultados de la validación de la selección de características HITON_MB . .	81
18.	Resultados de la validación de la selección de características HITON_MB al conjunto completo	82

Parte I

Objetivos

Este proyecto tiene los siguientes objetivos a cumplir.

1. En primer lugar queremos emplear la herramienta de relación causal de variables para establecer todas las posibles relaciones causa-consecuencia en el problema médico en [5], a partir de una base de datos obtenida del estudio previo.
2. En segundo lugar queremos encontrar una aplicación a los algoritmos de selección de características al mismo estudio. Se ha estado empleando de múltiples maneras, pero en este caso, lo aplicamos en un caso médico real dándole aún más relevancia.. Se ha demostrado su efectividad en múltiples ejemplos, al igual que haremos en este proyecto.

Dicho problema consiste en discriminar taquicardias ventriculares y supraventriculares a través de una serie de medidas tomadas tras una descarga de pulsos para controlar la arritmia.

Para llevar acabo este proyecto seguimos los siguientes pasos, que se exponen brevemente:

1. Inicialmente haremos un estudio sobre fisiología del corazón para entender el entorno en el que trabajaremos, visualizar los tipos de datos de los que disponemos, las clases a las que pertenecen nuestros datos, y finalmente poder comprender las posibles implicaciones de nuestros resultados.
2. A continuación escogeremos las posibles herramientas que podíamos utilizar para llevar a cabo el proyecto. Decidimos usar dos algoritmos diferentes para la selección de características, dos algoritmos para la relación causal de variables y SVM (Support Vector Machine ó Maquina de Vectores Soporte) a la hora de clasificar muestras para validar los resultados obtenidos por los algoritmos.

3. Más adelante, replicaremos el experimento recogido en el estudio “**Differentiation of Ventricular and Supraventricular Tachycardia based on the analysis of the first postpacing interval after sequential anti-tachycardia pacing in implanted cardioverter-defibrillator patients**” para comprobar que la SVM elegida era capaz de validar los resultados que obtendríamos más adelante.
4. A continuación, usaremos las herramientas seleccionadas con los datos médicos para obtener las características relevantes del estudio descrito arriba y buscar el árbol causal completo de las variables estudiadas, intentando averiguar que factores son los que causan las arritmias cardíacas y cuales son consecuencia de las mismas.
5. Finalmente, validaremos los datos obtenidos realizando un experimento de clasificación usando únicamente las variables obtenidas como resultado de la **selección causal de características**.

Nuestro objetivo primordial es el estudio de la **relación causal de variables**, ya que es la herramienta principal en el proyecto. Aunque nuestro proyecto consta de múltiples fases, en todas ellas hemos realizado como último paso una clasificación usando una máquina de vectores soporte para poder validar nuestros resultados con los obtenidos en otros estudios diferentes y ver si mantenían una lógica consistente.

1. Fisiología

Antes de comenzar con el proyecto como tal, necesitamos hacer un breve estudio sobre fisiología cardiaca. Es necesario para comprender el origen de nuestros datos, la finalidad de nuestro proyecto y el porqué de alguna de nuestras decisiones. Usamos la información encontrada en [8],[9],[14] [15],[16],[17].

Principalmente, la parte más interesante e importante para nosotros es la electrofisiología: las señales eléctricas que dan lugar a los latidos del corazón. Durante el latido se producen tres fases: la contracción de las aurículas pasando la sangre hacia el ventrículo; la contracción de los ventrículos enviando la sangre por todo el cuerpo y por último la relajación de todos los músculos que permiten la entrada de sangre nueva en el órgano. Éste sería el ciclo de un latido normal. Las arritmias cardiacas se producen cuando existe anomalías en este ciclo, ó en el periodo en el cual se ejecuta.

De todas las posibles arritmias, nos interesan principalmente las taquicardias, y de éstas, las ventriculares.

Las taquicardias ventriculares son el tipo más peligroso de taquicardia que hay. Se producen en el momento en el que el ventrículo aumenta su ritmo de bombeo llegando a alcanzar 240 latidos por minuto. El riesgo principal de la taquicardia ventricular es que puede desencadenar una fibrilación ventricular, en la cual no todas las fibras se contraen de golpe produciendo que no toda la sangre salga del ventrículo. Si ésto se diera y no fuera tratado de manera inmediata, existe un riesgo de muerte elevado.

Datos de los CDI Los dispositivos que se implantaron en los pacientes consistían, como visto en [5], en un sensor de > 5 mV para el electrodo ventricular con un umbral de $\frac{1,5V}{0,5ms}$ mientras que la salida para la terapia ATP (Anti-Tachycardia Pacing) se situaba en el vértice del ventrículo derecho, puesto que en el momento en el que se produce una taquicardia, la distancia entre el vértice ventricular derecho y el circuito taquicárdico ventricular formado por la arritmia, es menor que la distancia al atrio, y la corrección de un circuito ventricular

puede conseguirse con menos estímulos que cualquier arritmia supraventricular, lo cual nos interesa en este caso dada la gravedad de las taquicardias ventriculares.

A continuación vemos un poco más en detalle el corazón, su estructura y el ciclo cardiaco:

1.1. Estructura básica del corazón:

El corazón se localiza en el centro del tórax, entre los pulmones derecho e izquierdo, por detrás del esternón, entre el segundo y el quinto espacio intercostal izquierdo. Está situado de forma oblicua: aproximadamente dos tercios a la izquierda del plano medio y un tercio a la derecha. Es un órgano musculoso y cónico situado en la cavidad torácica, que funciona como una bomba, impulsando la sangre a todo el cuerpo. El corazón impulsa la sangre mediante los movimientos de sístole y diástole. Veremos un poco más en profundidad la estructura básica del corazón:

1.1.1. Estructura del corazón:

De dentro hacia fuera el corazón presenta las siguientes capas:

- El endocardio, una membrana de endotelio y tejido conectivo de revestimiento interno, que entra en contacto con la sangre. Incluye fibras elásticas y de colágeno, vasos sanguíneos y fibras musculares especializadas, las cuales se denominan Fibras de Purkinje. En su estructura encontramos las trabéculas carnosas, que dan resistencia para aumentar la contracción del corazón.
- El miocardio, el músculo cardíaco propiamente dicho; encargado de impulsar la sangre por el cuerpo mediante su contracción. Encontramos también en esta capa tejido conectivo, capilares sanguíneos, capilares linfáticos y fibras nerviosas.
- El epicardio, una capa fina serosa que envuelve al corazón llevando consigo capilares y fibras nerviosas. Esta capa se considera parte del pericardio seroso.

1.1.2. Cavidades cardíacas.

De tamaño poco más grande que un puño, el corazón está dividido en cuatro cavidades: dos superiores, llamadas aurículas, y dos inferiores, llamadas ventrículos. Los atrios reciben la sangre del sistema venoso, pasan a los ventrículos y desde ahí salen a la circulación arterial.

El atrio y el ventrículo derecho forman lo que clásicamente se denomina el corazón derecho. Recibe la sangre que proviene de todo el cuerpo, que desemboca en el atrio derecho a través de las venas cavas superior e inferior. Esta sangre, baja en oxígeno, llega al ventrículo derecho, desde donde es enviada a la circulación pulmonar por la arteria pulmonar. Su tamaño muscular es considerablemente menor al del ventrículo izquierdo, dado que la resistencia de la circulación pulmonar es menor que la sistémica, la fuerza que el ventrículo debe realizar es menor que la del izquierdo.

El atrio izquierdo y el ventrículo izquierdo forman el llamado corazón izquierdo. Recibe la sangre de la circulación pulmonar, que desemboca a través de las cuatro venas pulmonares a la porción superior de la aurícula izquierda. Esta sangre está oxigenada y proviene de los pulmones. El ventrículo izquierdo la envía por la arteria aorta para distribuirla por todo el organismo.

El tejido que separa el corazón derecho del izquierdo se denomina septo o tabique. Funcionalmente se divide en dos partes no separadas: la superior o tabique interatrial, y la inferior o tabique interventricular. Este último es especialmente importante, ya que por él discurre el fascículo de His, que permite llevar el impulso a las partes más bajas del corazón.

Válvulas cardíacas

Las válvulas cardíacas son las estructuras que separan unas cavidades de otras, evitando que exista reflujo retrógrado. Están situadas en torno a los orificios atrioventriculares (ó aurículo-ventriculares) y entre los ventrículos y las arterias de salida. Son las siguientes cuatro:

- La válvula tricúspide, que separa la aurícula derecha del ventrículo derecho.
- La válvula pulmonar, que separa el ventrículo derecho de la arteria pulmonar.

- La válvula mitral o bicúspide, que separa la aurícula izquierda del ventrículo izquierdo.
- La válvula aórtica, que separa el ventrículo izquierdo de la arteria aorta.

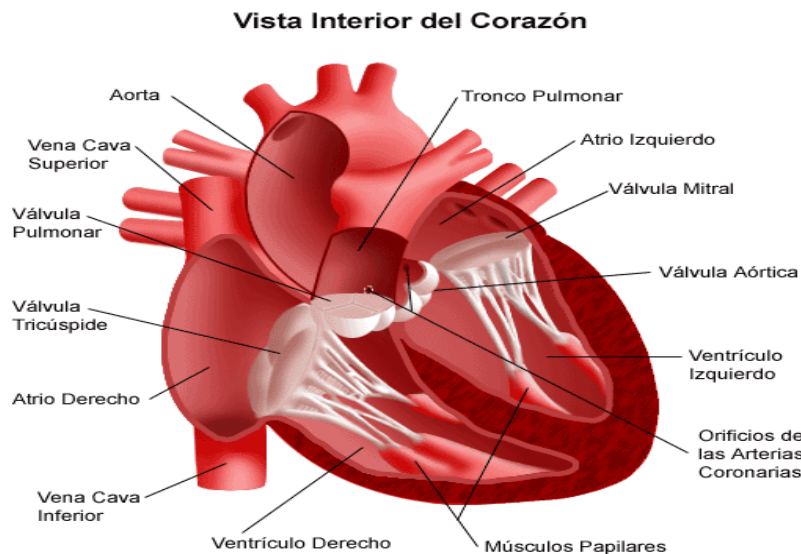


Figura 1: Estructura cardiaca del corazón

1.1.3. Ciclo Cardíaco:

Cada latido del corazón lleva consigo una secuencia de eventos que en conjunto forman el ciclo cardíaco, constando principalmente de tres etapas: sístole atrial, sístole ventricular y diástole. El ciclo cardíaco hace que el corazón alterne entre una contracción y una relajación aproximadamente 70 veces por minuto, es decir el ciclo cardíaco dura alrededor de 0,85 segundos.

Para que exista paso de sangre de una cavidad a otra del corazón, la presión de la cavidad impulsora ha de ser siempre mayor que la de la cavidad receptora.

Durante la sístole auricular, las aurículas se contraen y proyectan la sangre hacia los ventrículos. Si bien este paso de sangre es esencialmente pasivo y la contracción auricular participa

poco en condiciones de reposo, pero cobra importancia durante el ejercicio físico. Una vez que la sangre ha sido expulsada de las aurículas, las válvulas atrioventriculares entre las aurículas y los ventrículos se cierran. Esto evita el reflujo de sangre hacia las aurículas. El cierre de estas válvulas produce el sonido familiar del latido del corazón. Dura aproximadamente 0,1 segundos. En este momento el volumen ventricular es máximo, denominándose volumen de fin de diástole o telediastólico.

La sístole ventricular implica la contracción de los ventrículos expulsando la sangre hacia el aparato circulatorio. En esta fase se contrae primeramente la pared del ventrículo sin que haya paso de sangre porque hay que vencer la elevada presión de la aorta o de la arteria pulmonar; cuando esto se produzca tendrá lugar la eyección, la cual ocurre en dos fases, una rápida y otra lenta. Una vez que la sangre es expulsada, las dos válvulas sigmoideas, la válvula pulmonar en la derecha y la válvula aórtica en la izquierda, se cierran. Dura de manera aproximada 0,3 segundos.

Hay que decir que los ventrículos nunca se vacían del todo, quedando siempre sangre que forma el volumen de fin de sístole o telesistólico.

Por último la diástole es la relajación de todas las partes del corazón para permitir la llegada de nueva sangre. Dura aproximadamente 0,4 segundos.

En el proceso se pueden escuchar dos ruidos:

1. **Primer ruido cardiaco:** cierre de válvulas tricúspide y mitral.
2. **Segundo ruido cardiaco:** cierre de válvulas pulmonares y aortas.

Ambos ruidos se producen debido al cierre súbito de las válvulas, sin embargo no es el cierre lo que produce el ruido, sino la reverberación de la sangre adyacente y la vibración de las paredes del corazón y vasos cercanos. La propagación de esta vibración da como resultado la capacidad de auscultar dichos ruidos.

Este movimiento se produce de 70 a 80 veces por minuto.

La expulsión rítmica de la sangre provoca el pulso que se puede palpar en las arterias radiales, carótidas, femorales, etc.

Si se observa el tiempo de contracción y de relajación se verá que los atrios están en reposo alrededor de 0,7 segundos y los ventrículos unos 0,5 segundos. Ésto quiere decir que el corazón pasa más tiempo en estado de reposo que en trabajo.

En la fisiología del corazón cabe destacar que sus células se despolarizan por sí mismas dando lugar a un potencial de acción, que resulta en una contracción del músculo cardíaco. Por otra parte, las células del músculo cardíaco se "comunican" de manera que el potencial de acción se propaga por todas ellas, dando lugar a la contracción del corazón. El músculo del corazón jamás se tetaniza (los cardiomiocitos tienen alta refractariedad, es por eso que no hay tétanos).

1.2. Trastornos en la velocidad del ritmo cardíaco

Cuando el ritmo cardíaco está anormalmente acelerado (por encima de 100 latidos por minuto), con el paciente en reposo, se habla de **taquicardia**. Por el contrario, un ritmo cardíaco de menos de 60 latidos por minuto, constituye una bradicardia. Ni la taquicardia ni la bradicardia constituyen enfermedades por sí mismas, sino que son signos clínicos de alguna patología cardíaca subyacente. Algunas de las causas más frecuentes de taquicardia son el abuso de estimulantes, ciertos medicamentos que aceleran la frecuencia cardíaca, sobredosis de algunas drogas, inhalación de monóxido de carbono, etc. Entre las causas más frecuentes de bradicardia se encuentran los trastornos de la conducción, también llamados "bloqueos cardíacos".

Una taquicardia extrema puede hacer que los ventrículos se contraigan tan rápidamente que no alcancen a llenarse de sangre en cada ciclo cardíaco, con lo cual se produce shock y eventualmente la muerte.

Una bradicardia extrema hace que el volumen eyectado por el corazón sea normal, pero como los latidos son tan espaciados la cantidad que llega a los tejidos no es suficiente para

oxigenarlos, con los mismos resultados descritos anteriormente. Esto suele verse cuando no hay un funcionamiento adecuado del nódulo sinusal

En otras ocasiones, una parte del miocardio que normalmente no tendría actividad eléctrica propia genera un impulso que se propaga al resto del corazón y dispara un "latido extra", lo cual es conocido como extrasístole y se manifiesta con un síntoma muy conocido: la sensación de un vuelco en el corazón, con una pausa compensadora posterior. Cuando son pocas y aisladas no suelen requerir ningún tratamiento, pero siempre deberá ser el médico cardiólogo quien determine esto.

Resumiendo, tenemos los siguientes trastornos del ritmo cardíaco:

- Arritmia: toda irregularidad en el ritmo natural del corazón.
- Bradicardia: pulso muy lento, por debajo de 60 latidos por minuto.
- Taquicardia: pulso muy rápido, por encima de 100 latidos por minuto.
- Fibrilación: latidos rápidos no coordinados; son contracciones de fibras musculares cardíacas individuales.

Únicamente nos fijaremos en las taquicardias, que son los trastornos que nos interesan para este proyecto.

1.2.1. Taquicardia sinusal

La taquicardia sinusal es un aumento sostenido de la frecuencia cardíaca que excede de los límites normales de una persona. Suele producirse cuando se realiza un ejercicio físico continuado, aunque también puede producirse por exceso de cafeína, estrés, ó estado emocional alterado.

Los síntomas que presenta una persona con taquicardia sinusal varían desde poco más que el signo de aumento de la frecuencia cardíaca hasta un síndrome que puede impedir la actividad normal.

La palpitación es el síntoma más frecuente, aunque también puede presentarse intranquilidad, agitación, aprensión, ansiedad y molestias o dolor del pecho, según el umbral para los síntomas y la reactividad nerviosa de la persona. El diagnóstico guarda relación siempre al factor que la desencadena.

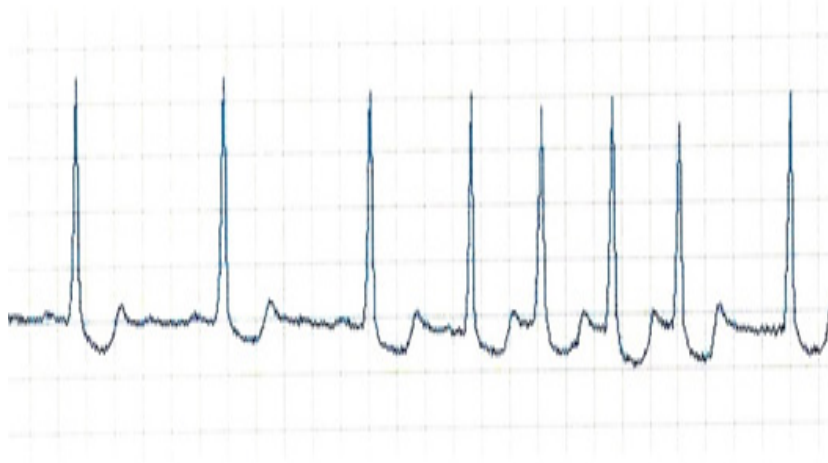


Figura 2: Ejemplo de taquicardia sinusal

La taquicardia sinusal no conlleva ningún riesgo, a menos que se produzca sin ningún tipo de relación con los factores explicados anteriormente, en cuyo caso debe buscarse la causa subyacente, puesto que puede indicar un mal funcionamiento del corazón ó un defecto congénito del mismo.

1.2.2. Taquicardia supraventricular

En cardiología, una taquicardia supraventricular es uno de los trastornos del ritmo cardíaco caracterizada por una frecuencia cardíaca acelerada cuya señal eléctrica se origina en el nódulo auriculoventricular ó bien en la aurícula cardíaca. Estas arritmias, por definición, se inician o se mantienen en el nódulo AV o en la aurícula, en contraste con las potencialmente letales taquicardias ventriculares, las cuales se originan en los ventrículos cardíacos, es decir, por debajo de las aurículas ó del nódulo AV. La frecuencia cardíaca normal es de unos 70 latidos

por minuto, mientras que en las taquicardias supraventricular, la frecuencia cardíaca varía entre 140-240 latidos por minuto.

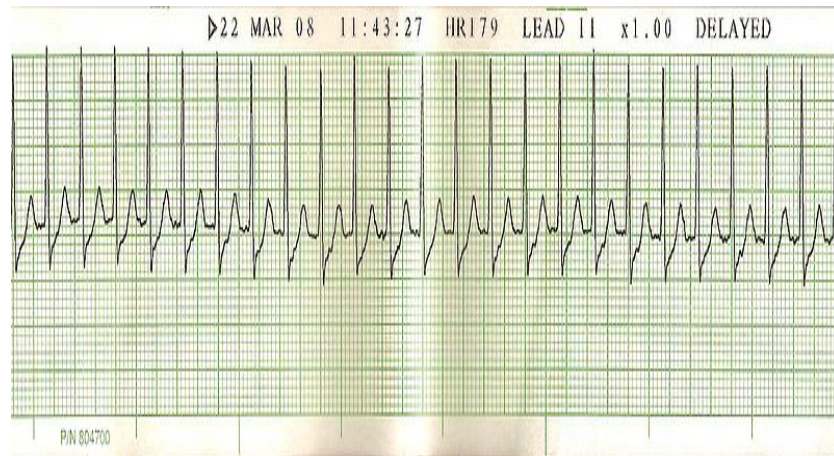


Figura 3: Ejemplo de Taquicardia supraventricular

Una vez que el episodio agudo ha terminado, el tratamiento continuo debe ser indicado para prevenir una recurrencia de la arritmia. Los pacientes que hayan tenido un solo episodio asilado ó episodios infrecuentes y mínimamente sintomáticos, por lo general no requieren tratamiento, solo observación ó seguimiento cardiológico regular.

Quienes tengan síntomas con más frecuencia ó síntomas debilitantes como consecuencia de episodios de taquicardia supraventricular, tienen la indicación de terapia preventiva medicamentosa como los agentes que bloquean al nódulo AV.

1.2.3. Taquicardia ventricular

Una taquicardia ventricular ó TV es uno de los trastornos del ritmo cardíaco caracterizada por una frecuencia cardíaca acelerada cuya señal eléctrica se origina en los ventrículos cardíacos, es decir, por debajo de las aurículas ó del nódulo AV. La frecuencia cardíaca normal es de unos 70 latidos por minuto, mientras que durante una TV el ritmo cardíaco puede llegar hasta 200-400 por minuto.

Muchas veces este ritmo no pone en peligro la vida del paciente, pero puede provocar síntomas tales como mareos, falta de aliento y palpitaciones (latidos fuertes del corazón). En algunas personas, las taquicardias ventriculares pueden dar lugar a un trastorno más grave llamado fibrilación ventricular (FV).

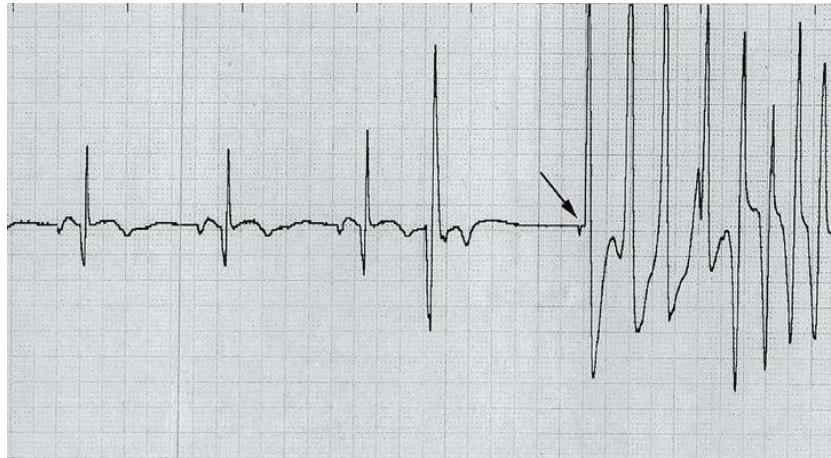


Figura 4: Ejemplo de taquicardia ventricular

El tratamiento de la taquicardia ventricular a largo plazo puede requerir el uso de medicamentos para la arritmia orales. Sin embargo, estos medicamentos pueden tener efectos secundarios severos y su uso actualmente está disminuyendo en favor de otros tratamientos.

En años recientes, se ha preferido un tratamiento para muchas taquicardias ventriculares crónicas que consiste en implantar un dispositivo llamado cardioversor desfibrilador implantable (CDI) en el tórax, como un marcapasos, y que se conecta al corazón por medio de cables.

Este dispositivo es programado por el médico para percibir la taquicardia ventricular cuando ésta se presente y para administrar un choque con el fin de abortarla. Asimismo, el dispositivo se puede programar para enviar una ráfaga de latidos rítmicos para interrumpir la taquicardia ventricular. Este tipo de taquicardia también puede requerir el uso de agentes antiarrítmicos concomitantes para prevenir el disparo del CDI. Este tratamiento es el que se ha usado en el estudio [5] y a partir del cual hemos obtenido nuestros datos para realizar el proyecto.

2. Selección causal de características

La determinación y la explotación de relaciones causales es algo básico a la hora de la toma de decisiones y de razonamiento. Un gran número de los algoritmos de selección de características emanan de aprendizaje máquina, sobre todo en las relaciones simétricas, y no buscan modelar los mecanismos: No pretenden revelar las relaciones causa efecto entre característica y objetivo. Ésto es justificable porque descubrir los mecanismos es innecesario a la hora de realizar buenas predicciones en un entorno puramente de observación. Se puede encontrar más información en los documentos [1], [2], [3] y [4].

2.1. Objetivo de la seleccion causal

El objetivo de determinar las relaciones causales es predecir las consecuencias de acciones dadas o de las manipulaciones. Aplicado de manera conjunta con la selección de características, tenemos la posibilidad de filtrar únicamente las variables relevantes para un objetivo concreto, y al mismo tiempo, encontrar la relación causa consecuencia que exista entre ellas, dándonos una visión del modelo para poder tomar decisiones con criterio para obtener el efecto deseado.

2.2. Fortalezas y debilidades

Una de las fortalezas de los algoritmos de selección de características es que las características relevantes se pueden discernir entre cientos de miles de distractores, con menos un centenar de muestras, en algunos dominios de problemas. En contraste, los modelos causales generalmente trabajan con pocas variables y un conocimiento casi perfecto de las distribuciones de las variables, lo que lleva implícito una abundancia de ejemplos de entrenamiento.

Los algoritmos de selección de características causales a veces realizan intervenciones en los datos que emplean; ésto es una manera de alterar la distribución “natural” de dichos datos,

lo cual viola el principio de datos distribuidos de manera idéntica e independiente, clásicamente aceptado en aprendizaje máquina. Bajo esta asunción, los datos de entrenamiento y evaluación se extraen de las mismas distribuciones.

Los algoritmos de selección de características poseen una debilidad adicional: seleccionan características en función de su efectividad a la hora de predecir el objetivo, sin tener en cuenta si esa capacidad predictiva proviene del sistema en cuestión ó de un artefacto experimental. Debido a ésto, es necesario conseguir robustez contra cambios en la distribución.

2.3. Posibles combinaciones

Existen oportunidades para una colaboración cruzada de ambos campos: El descubrimiento causal se beneficia de la selección de características para recortar la dimensionalidad por razones computacionales ó estadísticas (con el riesgo implícito de eliminar alguna característica relevante), y la selección de características se puede beneficiar del descubrimiento causal, acercándose a los mecanismos subyacentes y revelar una noción de relevancia mucho más refinada (a cambio de un mayor coste computacional).

Estamos hablando de descubrimiento causal, y a pesar de que la idea es intuitiva, vamos a definirla desde un punto de vista de ingeniería para poder explicar la herramienta que utilizamos a lo largo del proyecto.

2.4. Descripción

Desde el punto de vista del ingeniero, la causalidad es una idea muy orientada hacia un objetivo, que puede ser definida de manera sencilla como encontrar modos de actuar sobre un sistema para conseguir el resultado deseado. Introducimos el concepto de redes bayesianas, herramienta útil para entender y usar la idea de causalidad.

Las redes causales bayesianas proporcionan un entorno de trabajo conveniente para el razonamiento de causalidad entre variables aleatorias. Las redes causales bayesianas implementan una noción de orden causal y no modelan dependencias causales temporales en detalle

(aunque pueden extenderse para conseguirlo si así se desea). Aunque existen otros entornos, las redes bayesianas nos permiten ilustrar de manera sencilla las conexiones entre la selección de características y causalidad.

Recordamos que en un grafo acíclico dirigido (GAD), un nodo A es padre de B (B es hijo de A) si existe una flecha dirigida de A a B; A es antecesor de B (B es descendiente de A) si hay una flecha dirigida de A a B. Emplearemos la siguiente notación para las definiciones posteriores:

- Los términos “nodos” y “variables” se usan de manera indistinta.
- Las variables aleatorias se denotan con letra mayúscula X,Y,Z.
- Las realizaciones (valores) se denotan con letra minúscula x,y,z.
- Los conjuntos de variables ó de valores se denotan con letra mayúscula ó minúscula en negrita: $\mathbf{X} = [X_1, X_2, \dots, X_N]$ ó $\mathbf{x} = [x_1, x_2, \dots, x_N]$

Definimos formalmente una **Red Bayesiana Discreta**:

Supongamos que \mathbf{X} es un conjunto de variables aleatorias discretas y P una distribución de probabilidad sobre todas las realizaciones posibles de \mathbf{X} . Sea G un grafo acíclico dirigido y todos los nodos de G corresponden uno a uno con los miembros de \mathbf{X} . Necesitamos que por cada nodo $A \in \mathbf{X}$, A es probabilísticamente independiente de todos los no-descendientes de A, dados los padres de A (Condición de Markov). Llamaremos a la tupla $\{\mathbf{X}, G, P\}$ una red Bayesiana (discreta), ó equivalentemente una red de Creencia ó red probabilística.

A partir de esta definición podemos deducir la de **Red Bayesiana Causal**:

Red Bayesiana Causal. Una red Bayesiana causal es una red Bayesiana $\{\mathbf{X}, G, P\}$ con la semántica adicional que $(\forall A \in \mathbf{X})$ y $(\forall B \in \mathbf{X})$. Si hay una flecha desde A a B en G entonces A causa directamente B.

Es posible leer a partir de un grafo G si dos conjuntos de nodos A y B son independientes, condicionados a un tercer conjunto C : $A \perp_G B|C$. Sin embargo, en redes Bayesianas causales, la existencia de un camino dirigido entre dos nodos indica una relación causal.

Normalmente se asume que, en adición a la condición de Markov, que es parte de la definición de redes Bayesianas, se cumple otra condición conocida como “fidelidad”. La condición de fidelidad conlleva dependencias en la distribución desde el grafo mientras que la condición de Markov acarrea independencias en la distribución se recogen en el grafo. Juntas, las condiciones de Markov y fidelidad garantizan que la red Bayesiana será un mapa preciso de dependencias e independencias de la distribución representada.

Fidelidad de un GAD a una distribución: un grafo acíclico dirigido G es fiel a una distribución conjunta de probabilidad P sobre un conjunto de variables X , *si y sólo si* toda independencia presentada en P es supuesta por G y la condición de Markov, que es $\forall A \in X, \forall B \in X \text{ y } \forall C \subset X, A \perp_G B|C \Rightarrow A \perp B|C$.

Tanto la condición de Markov como la condición de fidelidad pueden ser especificadas de manera sencilla para las redes Bayesianas causales de la siguiente manera:

- **Condición causal de Markov (CMC):** en una red Bayesiana causal que satisfaga la CMC, cada nodo es independiente de sus no-efectos dadas sus causas inmediatas.
- **Condición de fidelidad de Markov (CFC):** una red Bayesiana causal es fiel si satisface la condición de fidelidad descrita anteriormente.

Las redes causales Bayesianas están totalmente definidas por su grafo y las probabilidades condicionales; éstas últimas nos pueden ser dadas por expertos ó entrenadas a partir de los datos (ó una combinación de ambas, como es nuestro caso). Una vez entrenada, la red Bayesiana puede usarse para computar cualquier probabilidad ó probabilidad condicional que implique cualquier subconjunto de variables. Dichos cálculos son conocidos como inferencia en redes Bayesianas. En los peores casos, la inferencia en redes Bayesianas es intrazable.

De todos modos, se han descrito una gran cantidad de algoritmos para inferencias exactas y aproximadas.

Por último vamos a ver los algoritmos de descubrimiento causal de una manera básica para explicar como funcionan.

Todos los métodos de aprendizaje máquina para el descubrimiento causal tienen dos componentes fundamentales: Un lenguaje para expresar modelos causales y una preferencia inductiva de ciertos tipos de modelos causales sobre otros. Esta breve presentación sobre algoritmos usa redes Bayesianas causales como el lenguaje para representar relaciones causales, y un tipo especial de distribuciones (distribuciones fieles) como la principal preferencia inductiva de estos algoritmos.

Asunciones hechas: asumimos que el conjunto de variables X a emplear es auto suficiente para caracterizar todas las dependencias causales e interés. La definición de suficiencia puede no ser clara, pero daremos un significado preciso en este caso.

Si asumimos que todas las variables que pueden ser seleccionadas a partir de un conjunto S de variables indicador que representan todos los eventos posibles en nuestro sistema bajo estudio, nuestra definición es la siguiente:

Suficiencia causal: un conjunto indicador de valores aleatorias $X = [X_1, X_2, \dots, X_N] \in S$ se dice que es causalmente autosuficiente si no hay conjunto de dos ó más variables en \mathbf{X} que posean una causa común en $S \setminus \mathbf{X}$.

Para inferir relaciones causales en un conjunto de variables \mathbf{X} , que no son causalmente suficientes es más difícil puesto que pueden existir dependencias entre las variables que no pueden ser explicadas causalmente desde las variables de X , pero a través de los confusores $S \setminus \mathbf{X}$. Ésto puede modelarse con variables “latentes” adicionales dando forma a factores desconocidos.

En los siguientes puntos, tendremos en cuenta las siguientes consideraciones:

- El conjunto de variables X es causalmente auto-suficiente.

- La máquina tiene acceso a un conjunto de entrenamiento suficientemente grande y a tests estadísticos fiables para determinar las independencias y las dependencias condicionales en la distribución original de donde se extraen las muestras de los datos.
- El proceso que genera los datos tiene una distribución $P(X,Y)$ que puede ser fielmente representado por la familia de modelos que tenemos en consideración (en este caso únicamente redes Bayesianas causales).
- El espacio de hipótesis es el espacio de todos los posibles modelos en consideración.

2.5. Un prototipo de algoritmo de descubrimiento causal

Enfatizamos la operación fundamental del algoritmo PC (técnicas de alta velocidad de barrido y detalles de implementación para simplificar la presentación).

Bajo las asunciones de descubrimiento causal [1] que hemos enumerado antes, este algoritmo es capaz de recuperar la estructura de una red Bayesiana que generó los datos, hasta una clase equivalente de Markov.

Suponemos que A , B y C con variables en \mathbf{X} y que \mathbf{V} es cualquier subconjunto de \mathbf{X} :

El algoritmo comienza con un grafo totalmente conectado sin orientar y consta de tres fases:

1. En primer lugar, encuentra los ejes dirigidos usando el criterio de que A comparte una unión directa con la variable B si y sólo si no existe un subconjunto de otras variables V que puede volverlas condicionalmente independientes ($A \perp B|V$).
2. Orientamos los ejes en trillizos colisionadores (del estilo $A \rightarrow C \leftarrow B$) usando el criterio de que hay ejes directos entre A , C y entre B , C , pero no entre A y B , entonces $A \rightarrow C \leftarrow B$, si y sólo si no hay un subconjunto V conteniendo C tal que $A \perp B|V$
3. Orientamos los ejes más lejanos con un método de propagación-limitación añadiendo orientaciones adicionales hasta que no se puede producir ninguna orientación más usando los dos criterios siguientes.

- a) Si $A \rightarrow B \rightarrow \dots \rightarrow C$ y $A - C$ (sino hay un eje no dirigido entre A y C) $A \rightarrow C$.
- b) Si $A \rightarrow B - C$ entonces $B \rightarrow C$.

Recientemente se han propuesto dos algoritmos de Manto de Markov, IAMB y HITON, que encuentran ejes directos no dirigidos hacia la variable objetivo, haciendo distinciones entre padres, hijos y esposas, proporcionando más información que una lista de características relevantes.

La diferencia con respecto al algoritmo PC es que éste comienza con un grafo totalmente conectado, mientras que HITON y IAMB comienzan con un grafo vacío. Además, los dos últimos utilizan métodos heurísticos para acelerar la búsqueda de grafos, que ha demostrado ser muy eficaz en la práctica. Incluyen la limitación de buscar los conjuntos condicionantes permitiendo la estimación de probabilidades condicionales y heurísticos de clasificación de variables que son candidatos a padres/hijos de Y.

De todas formas, los dos algoritmos usan heurísticos diferentes para reducir la búsqueda y realizar comprobaciones de sanidad. Estos algoritmos escalan bien para bases de datos muy grandes con cientos de miles de variables. En experimentos publicados se ha comprobado que HITON elimina falsos positivos que pueden haber sido encontrados por MMMB y se ha probado específicamente para la selección de características para clasificación con resultados muy prometedores.

Una gran novedad de los métodos locales es el sorteo de conectividad de grafos no uniforme. Una red puede no ser uniformemente densa. En un ámbito de aprendizaje global, si una región es particularmente densa, ese descubrimiento no puede ser realizado de manera rápida y cuando se aprende con una pequeña muestra, producirá muchos errores. Estos errores se propagan a las regiones remotas en la red. Al contrario que los métodos locales, que son rápidos y certeros en regiones menos densas. De ahí que tanto HITON como IAMB se comparan bien con el algoritmo estándar PC para descubrir redes Bayesianas completas.

Localizar la búsqueda para ejes dirigidos es muy deseable pero difícil computacionalmente. Una explicación de alto nivel es que cuando construimos los conjuntos de padres/hijos alrededor de Y de manera localizada, no incurrimos en falsos negativos y ocasionalmente omitiremos

variables X_i que no conectan con Y pero si con otras variables variables X_j que no son ni padres ni hijos de Y . Variables como X_i actúan como “variables ocultas” en el grado que concierne al criterio de independencia, abandonando falsos positivos. Resulta que:

1. La configuración en la cual sucede este problema es muy específica y en la práctica el número de falsos positivos es mínimo.
2. Los falsos positivos pueden ser detectados ejecutando un criterio de localización en la dirección opuesta.

Esto constituye la corrección de simetría de aprendizaje localizado de ejes dirigidos.

3. Clasificación mediante decisión máquina

Antes de comenzar la explicación sobre la decisión máquina debemos ver una serie de generalidades y de conceptos previos.

- **Clase:** todo dato que forme parte de nuestro problema, pertenece a un conjunto que denominaremos clase ó Hipótesis; lo denotaremos como H_i siendo i la clase a la que pertenece.
- **Muestra:** cada dato del cuál sabemos a que clase pertenece es conocido como muestra; lo denotamos como X^i siendo i el número de muestra.
- **Peso:** según la variable, tendrá una mayor ó menor influencia en el comportamiento de la muestra. Dicha influencia es el peso, y lo denotamos como w .

La decisión máquina consiste en diseñar una función decisora, a partir de una serie de muestras conocidas y de los pesos de cada una de las variables, que sea capaz de “decidir” a que clase pertenecen muestras desconocidas para nuestra máquina. Ésto es definir una función:

$$F_W(X) \gtrless_{D0}^{D1} U_{mbra}l$$

Si el valor obtenido una vez introducida la muestra es superior al umbral, decidimos que dicha muestra pertenece a la clase H_1 ; si es menor que el umbral la clasificamos como perteneciente a la clase H_0 .

Antes de continuar, debemos elegir la arquitectura de la máquina, que equivale a la forma de la función de decisión:

1. **Lineal:** $F_W(X) = w_0 + w_1x_1 + w_2x_2 + \dots + w_Nx_N$
2. **Cuadrática:** $F_W(X) = w_0 + w_1x_1 + w'_1x_1^2 + w_2x_2 + w'_2x_2^2 + \dots + w_Nx_N + w'_Nx_N^2$

3. **Senoidal:** $F_W(X) = w_0 + w_{1A}\text{sen}(w_{1B}x_1 + w_{1C}) + \dots + w_{NA}\text{sen}(w_{NB}x_N + w_{NC})$

Una vez elegida la arquitectura de la máquina, pasamos a ajustar los valores de los pesos (w_i). Para ello entrenamos la máquina pasándole muestras etiquetadas para los que ya conocemos la decisión correcta.

Hay que tener en cuenta que no debemos incorporar a la máquina todas las variables disponibles; evitaremos aquellas que sean irrelevantes (que no tengan relación con las hipótesis) ó que sean redundantes, aportando información que ya está incluida en otras. De ahí que apliquemos la decisión máquina a un problema médico que previamente ha sido sometido a un tratamiento de limpieza de datos incompletos, y a una selección de características.

Así mismo conviene usar en el entrenamiento un conjunto representativo de las muestras etiquetadas, a ser posible manteniendo la misma proporción de muestras entre las clases y el conjunto total para optimizar el clasificador.

Nuestro objetivo por lo general no será minimizar la probabilidad de error, sino minimizar los costes: definiremos un coste en función del tipo de error y le damos un valor. A partir de ahí elegimos la forma de la función F_W y a continuación entrenamos la máquina pasando muestras etiquetadas, ajustando los w para ir minimizar el coste.

Hay que tener cuidado a la hora de entrenar la máquina, porque podemos caer en dos situaciones problemáticas.

- **Sobreajuste:** entrenamos la máquina en exceso, de manera que la solución es demasiado particular, perdiendo capacidad de generalización para futuras aplicaciones. (Exceso de parámetros)
- **Subajuste:** No extraemos toda la información disponible del conjunto de muestras de entrenamiento, por tanto nuestra frontera no es óptima. Podríamos mejorar la solución pasándole más muestras ó un mayor número de parámetros sin perder generalidad. (Falta de parámetros)

Como nos interesa tener un tipo de función que sea capaz de aproximar cualquier curva, utilizamos polinomios. Son conocidos como aproximadores universales, ya que al aumentar el grado del mismo podemos aproximar cualquier curva.

Existe un inconveniente respecto a esta solución. Si el número de variables es demasiado alto, el número de parámetros para la función definida crece de manera muy rápida dando lugar a lo que es conocido como **explosión dimensional de los polinomios**. Si tenemos 14 variables, con un polinomio de grado 1 tendremos 15 parámetros, mientras que en un polinomio de grado 2 tendremos 124 parámetros.

Si se produjera ésto, la solución que obtendríamos sería muy buena, pero no tendría utilidad práctica debido al alto coste computacional necesario para alcanzarla y llegado el momento de aplicarla, las condiciones podrían haber cambiado llegando incluso a no tener alguna utilidad.

Parte II

Herramientas

En este apartado introducimos las herramientas que hemos empleado para desarrollar el proyecto, así como su funcionamiento y algunos ejemplos de aplicación para facilitar la comprensión de las mismas.

1. Árboles causales y Manto de Markov

Usadas para encontrar y estructurar relaciones de causa-efecto de un conjunto de variables. Seleccionamos dos tipos de algoritmos de causalidad y dos tipos de algoritmos de selección de características para cotejar resultados, tal como se enuncia en el artículo **'Casual feature Selection by Isabelle Guyon, Clopinet, California Constantin Aliferis, Vanderbilt University, Tennessee André Elissee, IBM Zürich, Switzerland March 2, 2007'** obteniendo en algunos casos diferentes resultados, los cuales pudimos validar usando clasificación máquina empleando la siguiente herramienta.

2. SVM: Máquina de Vectores Soporte

Son clasificadores usados para validar cada uno de los resultados obtenidos a través de la selección de características. Cada vez que obtenemos las características relevantes, las validamos con múltiples tipos de clasificadores: lineales, polinomiales y gaussianos, para ver cual de ellos se ajustaba mejor a nuestros datos obteniendo los resultados óptimos, haciendo una búsqueda de parámetros de coste, extensión de gaussianas, etc... en cada subapartado para encontrar la mejor máquina posible.

3. LIBSVM

Librería descargable de www.csie.ntu.edu.tw/~cjlin/libsvm/ implementada en MatLab para la aplicación de las Máquinas SVM. Tras una breve introducción, ilustramos con

una serie de ejemplos la variación de los resultados según vamos ajustando los parámetros principales.

Una vez presentadas nuestras herramientas, se proporciona una definición más formal y detallada sobre cada una de ellas.

4. Árboles Causales y Manto de Markov (Markov Blanket)

4.1. Manto de Markov

Definimos el Manto de Markov en una red bayesiana como el conjunto de nodos que protegen a un nodo objetivo de la influencia de otros nodos. En otras palabras, un Manto de Markov d-separa un nodo de otros nodos que no pertenecen al Manto de Markov.

Formalmente hablando, si $\mathbf{X} \cup Y$ ($Y \notin \mathbf{X}$) es el conjunto de todas las variables en consideración, y \mathbf{V} un subconjunto de \mathbf{X} . Denotamos por “ \setminus ” el conjunto diferencia.

Manto de Markov. Un subconjunto \mathbf{M} de \mathbf{X} es un Manto de Markov de $Y \iff$ para cualquier subconjunto V de \mathbf{X} , es independiente de $V \setminus \mathbf{M}$ dado \mathbf{M} (p.e $Y \perp V \setminus \mathbf{M} \mid \mathbf{M}$, esto es $P(Y, V \setminus \mathbf{M} \mid \mathbf{M}) = P(Y \mid \mathbf{M}) P(V \setminus \mathbf{M} \mid \mathbf{M})$ ó para $P(V \setminus \mathbf{M} \mid \mathbf{M}) > 0$, $P(Y \mid V \setminus \mathbf{M}, \mathbf{M}) = P(Y \mid \mathbf{M})$).

Los Mantos de Markov por lo general no son únicos y pueden variar en tamaño. Sin embargo, si una red bayesiana cumple las condiciones de Markov y de fidelidad, tendrá un único Manto de Markov que incluirá sus causas directas (padres), los efectos directos (hijos), y las causas directas de efectos directos (esposas). Hay que considerar lo siguiente: no incluye las consecuencias directas de causas directas (hermanos), ni las causas directas de las causas directas (abuelos). Para entender la idea detrás los Mantos de Markov, vamos a estudiar el ejemplo que ponemos a continuación:

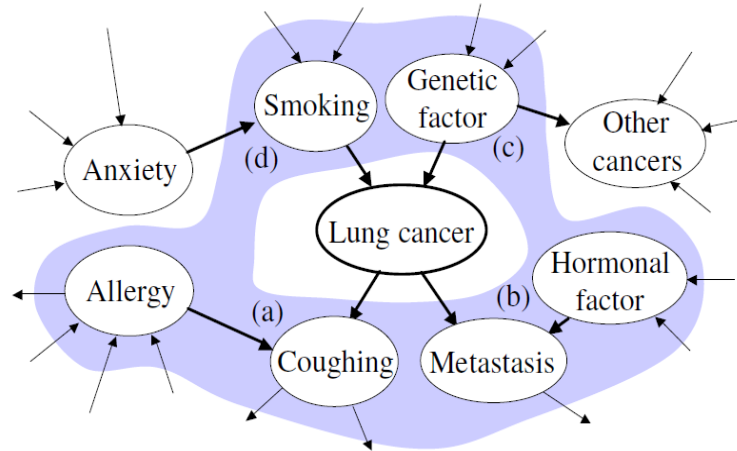


Figura 5: Manto de Markov aplicado a la variable Lung Cancer
Imagen obtenida de [1]

En este caso, el Manto de Markov de la variable “Lung Cancer” lo componen las variables “Smoking” y “Genetic Factor” como padres. “Metastasis” y “Coughing” como hijos y, por último, “Hormonal Factor” y “Allergy” son esposas. Vemos que quiere decir ésto de manera más detallada:

- **Causas directas (padres):** Una vez que todas las causas se han dado, una causa indirecta (“anxiety”) no trae ningún tipo de información adicional. Por ejemplo, en la figura 6-d un aumento en la ansiedad eventualmente incrementará “smoking” pero no influencia directamente “Lung Cancer”, así que nos es suficiente con usar “smoking” como predictor, no necesitamos saber acerca de “anxiety”. De manera similar, cualquier consecuencia de causa directa (como “Other cancers” en la figura 6-c, que es consecuencia de “Genetic factor”) sólo proporciona evidencia indirecta, pero no información adicional una vez que la causa directa “Genetic factor” es conocida. Las causas directas en distribuciones fieles son predictivas individualmente, pero puede ser necesario conocerlas de manera conjunta para que puedan ser predictoras.

- **Efectos directos (hijos) y causas directas de efectos directos (esposas):** En distribuciones fieles, los efectos directos son siempre predictores del objetivo. Pero su potencia predictora puede ser mejorada sabiendo otras causas posibles de estos efectos directos. Por ejemplo, en la figura 6-a “allergy” puede causar “coughing” independientemente de que exista “Lung Cancer”. Es importante conocer cualquier tipo de “allergy”, que eventualmente podría ayudarnos a descartar “coughing” como resultado de “Lung Cancer”. Las esposas, que no tienen un camino directo al objetivo (como “allergy”) y no son predictores individualmente del objetivo “Lung Cancer”, necesitan tener un hijo común (“coughing”) para poder llegar a ser predictoras. De todos modos, en distribuciones infieles, los hijos no son necesariamente predictores sin la ayuda de las esposas.

Si interpretamos la noción del Manto de Markov en distribuciones fieles en términos de relevancia de características, obtenemos lo siguiente:

1. **Irrelevancia:** una característica es irrelevante si está desconectada de Y (objetivo) en el grafo.
2. **Relevancia fuerte:** las características fuertemente relevantes forman el Manto de Markov M de Y .
3. **Relevancia débil:** las características que tienen un camino conectado a Y , pero no pertenecen a M , son débilmente relevantes.

4.2. HITON MB e IAMB

Estos algoritmos locales de descubrimiento causal nacieron como alternativa a la aplicación del algoritmo PC en caso de darse una serie de circunstancias excepcionales. Cuando el conjunto de datos tiene decenas ó cientos de miles de variables, ó cuando al menos algunos de ellos están estrechamente interconectados, la aplicación de algoritmos estándar de descubrimiento causal que puedan extraer la red completa resulta poco práctica. Para estos casos se pueden usar este tipo de algoritmos, que se centran en entender la estructura de la red únicamente en la región inmediatamente vecina de Y .

Según [1] y [2], tanto HITON MB como IAMB encuentran nexos directos a la variable objetivo Y (sin dirección) haciendo una distinción entre padres, hijos y esposas, por lo que proporcionan más información que una mera lista de características fuertemente relevantes.

Aparte de limitar la búsqueda a encontrar nexos a la variable objetivo y buscar sus esposas, estos algoritmos usan un número de heurísticos para acelerar la búsqueda que han resultado ser muy eficientes en la práctica. Permiten limitar la búsqueda a conjuntos condicionales de tamaño que permitan la estimación firme de probabilidades condicionales, y heurísticos de clasificación de variables que son candidatas a ser padres ó hijos de la variable objetivo.

Estos algoritmos escalan muy bien para bases de datos grandes, del orden de 10^5 variables. En experimentos previos se ha mostrado que ambos experimentos funcionan bien, pero con la salvedad de que HITON elimina algunos falsos positivos que podrían ser considerados por IAMB, siendo examinado para la selección de características para clasificación con resultados prometedores.

Otra ventaja de los métodos locales es la capacidad de sortear la conectividad no-uniforme de los grafo, siendo rápidos y precisos en regiones poco densas, evitando la propagación de errores a regiones remotas de la red. De este modo, HITON e IAMB comparan bien respecto al algoritmo estándar PC para el descubrimiento de redes Bayesianas completas.

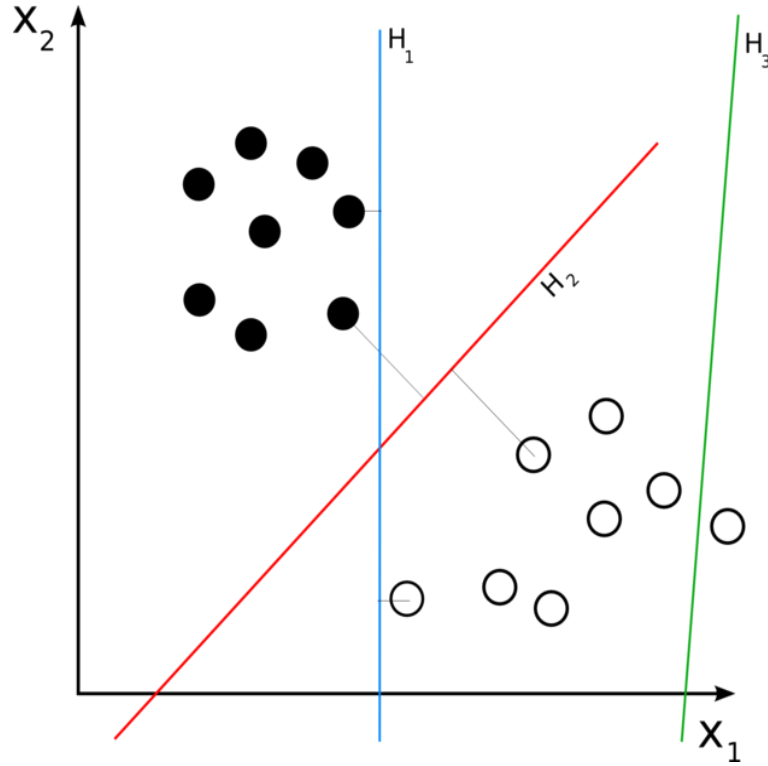
5. SVM: Support Vector Machine

Definición

Las SVM (máquinas de vectores soporte) pertenecen a una familia de clasificadores lineales generalizados. Pueden llegar a considerarse un caso especial de la regularización de Tikhonov. Una propiedad especial es que simultáneamente minimizan el error empírico de clasificación y maximizan el margen geométrico; por tanto son también conocidos como clasificadores de margen máximo. Ver [19].

Las SVM consisten en un conjunto de métodos de aprendizaje máquina supervisados los cuáles se emplean principalmente para clasificación y para regresión. Si consideramos los datos de entrada como conjuntos de vectores en un espacio n -dimensional, una SVM construirá un hiperplano separador en dicho espacio, uno que maximizará la distancia entre los dos conjuntos de datos. Para calcular dicho margen, se crean dos hiperplanos paralelos, uno a cada lado del hiperplano frontera separador, y “se empujan” hacia los dos conjuntos de datos hasta llegar a la primera muestra del mismo. Intuitivamente, el hiperplano consigue una mejor separación a mayor distancia de la vecindad de ambos conjuntos; puesto que generalmente, cuanto mayor sea el margen, menor será el error de generalización del clasificador.

La clasificación de los datos es una tarea muy común en aprendizaje máquina. Supongamos que se nos proporcionan unos datos, pertenecientes a dos posibles clases; y que nuestro objetivo es que dado un nuevo dato, poder decidir a que clase pertenece. En el caso de una máquina de vectores soporte, cada dato es visto como un vector p -dimensional (*una lista de p números*) y queremos saber si podemos separar dichos puntos con un hiperplano que tenga como máximo $p-1$ dimensiones. Ésto es conocido como un **clasificador lineal**. Existen un gran número de hiperplanos capaces de clasificar nuestros datos de manera correcta, pero nos interesa que exista una distancia máxima entre los dos conjuntos. Por tanto, nos interesa un hiperplano tal que la distancia entre los dos puntos más cercanos de ambos conjuntos sea máxima. Si dicha frontera existe, obviamente es la que nos interesa, y es conocida como el *hiperplano de margen máximo* y, en el caso de clasificación, como *clasificador de margen máximo*.



Podemos observar como en este caso tenemos 3 fronteras candidatas: H_3 no separa los dos conjuntos así que la descartamos. H_1 separa las dos nubes de muestras, pero con un margen mínimo. H_2 separa los dos conjuntos con el margen máximo posible, por tanto H_2 es la frontera óptima (clasificador de máximo margen).

Figura 6: Ejemplo de fronteras de clasificación.
Imagen obtenida de 'The support vector machine under test'.

Damos ahora una definición más formal de la SVM:

Supongamos que tenemos un conjunto de puntos de la forma:

$$D = \{(x_i, c_i) / x_i \in \mathbb{R}^p, c_i \in \{-1, 1\}\}_{i=1}^n$$

donde c_i es $+1$ ó -1 , indicando a que clase pertenece la muestra X_i . Cada X_i es un vector real p -dimensional. Queremos encontrar el hiperplano de máximo margen que dividirá los puntos

que tienen valor $c_i = 1$ de aquellos con valor $c_i = -1$. Cualquier hiperplano puede ser escrito como el conjunto de puntos \mathbf{X} que cumplen:

$$\mathbf{w} \bullet \mathbf{x}_i - b = 0$$

donde \bullet denota el producto de ambos términos. El vector \mathbf{w} es un vector normal: es perpendicular al hiperplano frontera. El parámetro $\frac{b}{\|\mathbf{w}\|}$ determina la compensación del hiperplano desde el origen a lo largo del vector normal \mathbf{W} . Queremos escoger tanto \mathbf{w} como \mathbf{b} para maximizar el margen ó la distancia entre hiperplanos paralelos que estén tan lejos entre sí como sea posible manteniendo separados las dos clases de datos. Estos hiperplanos pueden ser descritos (en este ejemplo) de la siguiente forma:

$$\mathbf{w} \bullet \mathbf{x}_i - b \geq 1 \quad \text{para } \mathbf{x}_i \text{ (Hiperplano paralelo 1)}$$

$$\mathbf{w} \bullet \mathbf{x}_i - b \leq -1 \quad \text{para } \mathbf{x}_i \text{ (Hiperplano paralelo 2)}$$

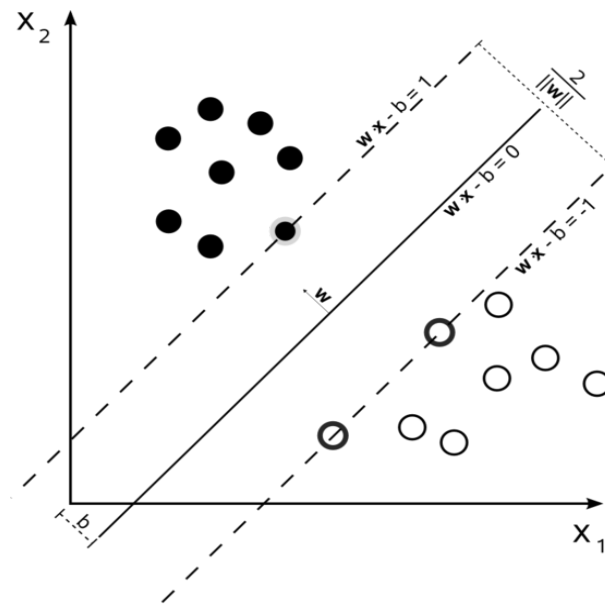


Figura 7: Ejemplo de dos hiperplanos paralelos al hiperplano frontera de margen máximo.
Imagen obtenida de 'The support vector machine under test'

Hay que destacar que si los datos de entrenamiento son linealmente separables, podemos seleccionar los dos hiperplanos del margen de modo que no haya puntos entre ellos e intentar maximizar la distancia mutua. Usando nociones de geometría, vemos que la distancia entre estos dos hiperplanos es $\frac{2}{\|w\|}$, así que queremos minimizar $\|w\|$. Puesto que tenemos que prevenir que haya datos que caigan dentro de dicho margen, podemos añadir la siguiente limitación:

$$\forall i \text{ debe cumplirse que: } c_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1. \text{ . (para todo } 1 \leq i \leq n)$$

Juntando todo esto, podemos llegar al problema de optimización:

Minimizar (en \mathbf{w} , b) $\|w\|$ propenso $\forall i \in (1, \dots, n)$

$$c_i(\mathbf{w} \bullet \mathbf{x}_i - b) \geq 1.$$

5.1. Forma primal

El problema de optimización presentado en el apartado anterior es difícil de resolver, porque depende de $\|w\|$, la norma de \mathbf{w} , que lleva implícita una raíz cuadrada. Afortunadamente, es posible alterar la ecuación sustituyendo $\|w\|$ por $\frac{1}{2}\|w\|^2$ sin cambiar la solución (el mínimo del original y de la ecuación modificada tienen el mismo \mathbf{w} y b). Éste es el problema de optimización de programación cuadrática. Más claramente:

Minimizar (en \mathbf{w} , b) $\frac{1}{2}\|w\|^2$ propenso (para cualquier $i=1, \dots, n$) :

$$c_i(\mathbf{w} \bullet \mathbf{x}_i - b) \geq 1.$$

El factor de $\frac{1}{2}$ se usa por convenio matemático. Este problema puede resolverse a través de técnicas de programación cuadrática estándar.

5.2. Forma dual.

Escribiendo la regla de clasificación en su forma dual no restrictiva revela que el hiperplano de margen máximo, y por tanto, la tarea de clasificación, es solamente una función de vectores soporte, mientras que los datos de entrenamiento permanecen en los márgenes.

Nos interesa maximizar (en α_i) la siguiente función:

$$\sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j c_i c_j \mathbf{x}_i^T \mathbf{x}_j \quad (1)$$

Los términos α_i constituyen una representación dual del vector de pesos en términos del conjunto de entrenamiento:

$$\mathbf{w} = \sum_i \alpha_i c_i \mathbf{x}_i. \quad (2)$$

Quedando como resultado la siguiente función de decisión:

$$\text{sgn}\left(\sum_{i=1}^l y_i \alpha_i K(x_i, x) + b\right)$$

5.3. Hiperplanos sesgados e insesgados

Por razones de simplificación, a veces se necesita que el hiperplano pase a través del origen del sistema de coordenadas. Tales hiperplanos son conocidos como insesgados, mientras que hiperplanos generales que no pasan necesariamente por el origen de coordenadas se les denomina sesgados. Un hiperplano insesgado puede ser válido por el parámetro $b = 0$ en el problema de optimización fundamental. Su correspondencia dual es idéntica a la explicada anteriormente sin la restricción de igualdad.

$$\sum_{i=1}^n \alpha_i c_i = 0. \quad (1)$$

5.4. Extensiones a la SVM lineal

En 1995, se sugirió la idea de margen máximo para ejemplos sin etiqueta. Si no existe un hiperplano que pueda separar los ejemplos de “Si” y “No”, el método de margen suave escogerá un hiperplano que separe los datos tan limpiamente como sea posible, mientras que maximizará la distancia al dato separado más cercano. Éste método introduce las variables conflictivas ξ_i , que miden el grado de error de clasificación del dato \mathbf{x}_i .

$$c_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1 - \xi_i \quad (1 \leq i \leq n.) \quad (2)$$

La función objetivo es entonces incrementada por una función que penaliza todo ξ_i que no sea cero, y la optimización se convierte en una búsqueda de equilibrio entre un margen amplio y una penalización por error leve. Si la función de penalización es lineal, el problema de optimización pasa a ser el siguiente:

$$\operatorname{argmin}_W \quad \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i \quad \text{restringido a } c_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1 - \xi_i \quad 1 \leq i \leq n. \quad (3)$$

Esta restricción en (2) junto con el objetivo de minimizar $\|\mathbf{w}\|$ puede resolverse aplicando multiplicadores de Lagrange. La ventaja fundamental de una función penalizadora lineal es que las variables conflictivas del problema dual, con la constante C apareciendo únicamente como una restricción adicional a los multiplicadores de Lagrange.

Las funciones penalizadoras no lineales se han usado, particularmente para reducir el efecto de ‘outliers’ en los clasificadores, pero a menos que se tenga cuidado, el problema pasa a ser no convexo, y por tanto aumenta tremendamente la dificultad de encontrar una solución global al perder las condiciones necesarias para hacer la aproximación simplista.

El algoritmo de hiperplano óptimo original era un clasificador lineal. No obstante, más adelante se sugirió un modo de crear clasificadores no lineales aplicando el truco de los kernels. El algoritmo resultante es formalmente similar, excepto que el producto escalar es reemplazado por una función de kernel no lineal. Ésto permite al algoritmo encajar el hiperplano de margen máximo en el espacio de características transformado. La transformación puede no ser lineal y el espacio transformado ser de una dimensión alta. De este modo, aunque el clasificador es un hiperplano en un espacio de características de dimensión alta, puede ser n-lineal en el espacio original de entrada.

Si el kernel usado es una función gaussiana de base radial, el espacio característico correspondiente es un espacio de Hilbert de dimensión infinita. Los clasificadores de margen máximo están bien regularizados, así que la dimensión infinita no echa a perder los resultados. Algunos kernels comunes incluyen:

1. Lineal: $k(\mathbf{x}, \mathbf{x}') = c(\mathbf{x} \cdot \mathbf{x}')$
2. Polinomial: $k(\mathbf{x}, \mathbf{x}') = (\mathbf{x} \cdot \mathbf{x}')^d + c$
3. Función de base radial: $k(\mathbf{x}, \mathbf{x}') = e^{(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2)}$, para $\gamma > 0$
4. Función gaussiana de base radial: $k(\mathbf{x}, \mathbf{x}') = e^{\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right)}$.
5. Función sigmoide: $k(\mathbf{x}, \mathbf{x}') = \tanh(\kappa \mathbf{x} \cdot \mathbf{x}' + c)$. con ciertas limitaciones: $c < 0$ y $\kappa > 0$

La principal contra a la hora de usar una SVM para la clasificación es la siguiente:

Las SVM se pueden aplicar directamente únicamente para tareas binarias, Por tanto, se deben aplicar algoritmos que reduzcan los problemas multiclase a problemas binarios.

5.5. SVM Multiclase

Las SVM multiclase pretenden asignar etiquetas a ejemplos usando máquinas de soporte vectorial, dónde las etiquetas son extraídas de un conjunto finito de varios elementos. La aproximación dominante para hacer ésto es reducir el problema singular multiclase en múltiples

problemas binarios. Cada uno de esos problemas da como resultado un clasificador binario, que se asume produce una función de salida que da valores relativamente grandes para ejemplos de la clase positiva y valores relativamente pequeños para ejemplos que pertenecen a la clase negativa.

Existen dos métodos para construir tales clasificadores binarios. Cada clasificador puede diferenciar entre:

1. Una de las etiquetas frente al resto.
2. Entre cada pareja de clases.

La clasificación de nuevas muestras para el caso 1 se realiza a través de la estrategia del ganador-mantiene-todo, a partir del cuál el clasificador con la función de salida más alta asigna la clase. La clasificación del caso 2 se realiza a partir de la estrategia del mayor número de votos en la cual cada clasificador asigna cada muestra a una de las dos clases candidatas. Después, el voto por la clase asignada se incrementa por un voto, y finalmente la clase con mayor número de votos determina la clasificación de dicha muestra.

5.5.1. Implementación

Los parámetros del hiperplano de margen máximo se derivan de la resolución de la optimización. Existen diferentes algoritmos especializados para la rápida resolución del problema de programación cuadrática (PC) que surge de las SVM, que en su mayor parte dependen de métodos heurísticos para dividir el problema en trozos más pequeños y manejables. Un método común para la resolución del problema PC es el algoritmo de Optimización Mínima Secuencial de Platt, que divide el problema a subproblemas de dos dimensiones que pueden ser resueltos analíticamente, eliminando la necesidad de utilizar un algoritmo de optimización numérico.

6. LIBSVM

LIBSVM es un software integrado que permite la clasificación por vectores soporte (C-SVC, ν -SVC), regresión (ϵ -SVR, ν -SVR) y estimación de distribuciones (SVM uniclase). Soporta clasificación multiclase. Desarrollado por Chih-Chung Chang y Chih-Jen Lin, busca permitir a todos los usuarios la capacidad de usar SVM como una herramienta de una manera sencilla. LIBSVM proporciona una interfaz sencilla a partir de la cual los usuarios pueden enlazarlo con sus propios programas, en nuestro caso MatLab.

A la hora de ejecutar el módulo tenemos una serie de opciones a configurar según la función que queramos realizar. A continuación detallamos los múltiples parámetros que hemos utilizado para nuestro proyecto, con su significado:

1. **-s tipo de svm** : configura el tipo de SVM (por defecto 0).
0 – C-SVC (Clasificación).
2. **-t tipo de kernel** : Configura la función de kernel a utilizar sobre las muestras (por defecto 2).
0 – lineal $u' * v$.
1 – polinomial $(\gamma * u' * v + coef0)^{grado}$.
2 – función de base radial $e^{(-\gamma * |u-v|)^2}$.
3. **-d grado** : Configura el grado de la función de kernel (por defecto 3).
4. **-g γ** : Fija γ en la función de kernel (por defecto $\frac{1}{n^o \text{ de características}}$).
5. **-r coef0** : Fija el valor del coef0 de la función de kernel (por defecto 0).
6. **-c coste** : Fija el valor de C de C-SVC, ϵ -SVR, y ν -SVR (por defecto 1).
7. **-n ν** : Fija el parámetro ν de ν -SVC, SVM uniclase, y ν -SVR (por defecto 0.5)).

8. **-wi peso:** Fija el parámetro de coste C de la clase i a $\text{peso}_{\text{clase}_i} * C$, para C-SVC (por defecto 1).

En nuestro problema, realizamos un ajuste adicional debido la naturaleza médica de los datos. A diferencia de otros problemas, en nuestro caso no tenemos una simetría de errores. No es lo mismo encontrar una Taquicardia Ventricular cuando no lo es y viceversa. Por ello utilizamos unos costes ponderados en función de la cantidad de Taquicardias Ventriculares y Supraventriculares que teníamos en nuestros datos.

Para hallar dicho factor, dividíamos el número de muestras pertenecientes a una de las dos clases, entre el número de muestras de la otra clase, hallando la proporción entre ambas clases optimizando nuestros resultados. Este “factor de ajuste” de cada clase lo multiplicamos por el valor de Coste de error general, ajustando mejor la frontera de decisión para clasificar nuestros datos, para evaluar nuestros resultados.

Vemos diferentes ejemplos de como el valor de coste y de gamma afecta a las fronteras de decisión. Para ésto empleamos el applet que se encuentra en: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.

Ejemplos de aplicación de la SVM

Tenemos un problema binario: La clase 1 está representada por los puntos azules, mientras que la clase 0 está representada por los puntos magenta. La situación inicial del problema es la siguiente:

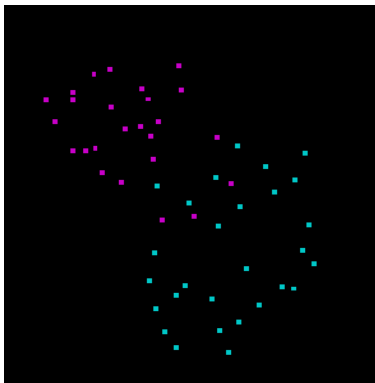


Figura 8: Situación inicial del ejemplo SVM.

Tenemos dos conjuntos principales de muestras, pero algunas de ellas están entremezcladas. Ésta será la zona conflictiva en la que nuestra SVM variará principalmente la medida que vayamos modificando parámetros.

Queremos ver cual sería la frontera de clasificación óptima en esta situación, suponiendo que el coste sea simétrico para ambos tipos de errores. Obtenemos lo siguiente:

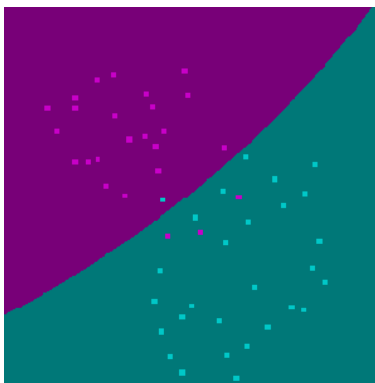


Figura 9: Frontera de decisión óptima en caso simétrico

Observamos como la frontera se traza de manera que la probabilidad de error sea lo menor posible para ambas clases. Estamos usando un kernel gaussiano, y un valor de coste de 100 para ambas clases.

Para trazar esta frontera, usamos como referencia las muestras más cercanas de cada conjunto al límite del otro, y usamos unos vectores soporte para ir trazando la curva.

A partir de este punto, vamos a ir modificando desde el valor de coste, el ancho de las gaussianas (γ), el valor de peso de cada clase, es decir, no asumiremos que el problema es simétrico, e iremos observando como va cambiando la frontera a medida que realizamos estas modificaciones.

Si modificamos la simetría de los costes obtenemos la siguiente frontera, usamos una SVM en modo clasificación con coste 200 y un valor de w_1 de 10 para ponderar nuestra clase 1.



Figura 10: Frontera de decisión ponderada para la clase 1

Vemos como la frontera se modifica curvándose para intentar evitar todos los errores de la clase 1, puesto que es mucho más perjudicial clasificar una muestra azul mal que múltiples muestras magenta.

Ahora vemos el contrario. Ponderamos el peso de error de la clase 0 y vemos como cambia la frontera:

Usamos una SVM en modo clasificación con coste 200 y un valor de w_0 de 10 para ponderar nuestra clase 0.

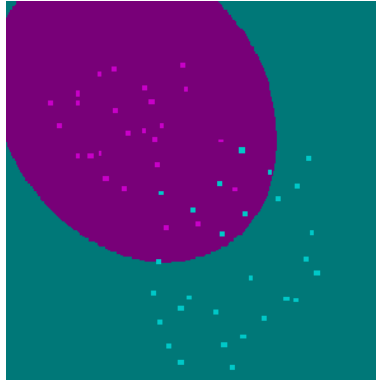


Figura 11: Frontera de decisión ponderada para la clase 0.

Observamos que sucede lo opuesto al caso anterior, la frontera de decisión se cierra alrededor de las muestras magenta, haciendo todo lo posible para evitar una mala clasificación de este conjunto debido al alto coste del error.

Ahora modificaremos la anchura de las gaussianas de las muestras, manteniendo un coste simétrico, y vemos como se modifica la frontera de decisión. Usamos una SVM en modo clasificación con coste 100 y un valor de γ de 2.

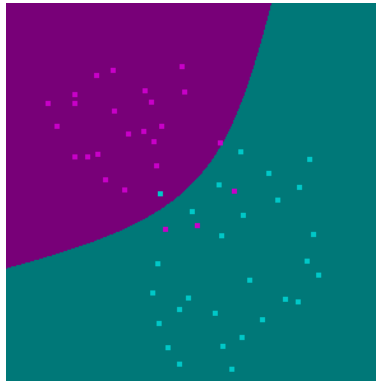


Figura 12: Frontera de decisión con factor de extensión de gaussianas y coste simétrico.

Observamos que tiene parecido con el caso de coste simétrico, con la diferencia de que es más curvada. Ésto se debe a que al añadir una extensión a las muestras, existe cierta aleatoriedad

añadida que se tiene en cuenta aumentando el número de vectores soporte que componen la frontera allí donde hay muestras de ambos conjuntos cerca.

Por último, vamos a ver que sucede si combinamos todos los factores que hemos estudiado por separado: usaremos costes no simétricos, ponderando cada clase por separado y con un valor de extensión de 2.5.

Usamos una SVM modo clasificación con coste de 200, un valor de g de 2.5, peso de la clase 1 de 1.5 y peso de la clase 0 de 2.



Figura 13: Frontera de decisión con coste, ponderación y extensión de gaussianas.

Podemos observar como la frontera se cierra ligeramente alrededor de las muestras magenta, debido a que hay un mayor coste por error en la clase 0, aunque no es demasiado elevado. Del mismo modo, observamos el comportamiento “oscilatorio” en la zona del vértice de la curva, donde se acumulan muestras de ambas clases cruzadas y la frontera va realizando ligeras correcciones para ajustarse lo máximo posible.

Por último mostramos como quedaría la frontera en caso de llegar a un sobreajuste, dando los mismos parámetros que en el ejemplo anterior, pero aumentando hasta 10000 el valor de coste.



Figura 14: Frontera de decisión sobreajustada

Aquí podemos observar el caso extremo del ajuste. Ésto no debe suceder nunca. Al haber elevado tanto el coste por error, la frontera se curva sobre nuestras muestras de tal manera que ha perdido generalidad. Se ha ajustado de tal forma a este problema en particular que en otras distribuciones resulta inútil.

Además, existe una gran complejidad a la hora de trazar esta frontera. Ello implica una carga computacional excesiva, un gran número de vectores soporte y un retardo a la hora de proporcionar la frontera; que en un caso real hará que sea inútil, puesto que no servirá en el momento en el que cambien las condiciones del problema.

6.1. Selección de parámetros

LIBSVM proporciona una herramienta de selección usando un kernel de función de base radial: validación cruzada a través de cuadrículas paralelas. Actualmente sólo se da soporte a dos parámetros: C y γ . Pueden modificarse de manera sencilla para otros kernels, ya sea lineal ó polinomial, ó incluso para la regresión por vectores soporte.

Para problemas de tamaño medio, la validación cruzada puede ser la forma más fiable para la selección de parámetros.

1. El conjunto de entrenamiento se divide en diferentes carpetas.

2. A continuación una carpeta es considerada como el conjunto de validación, mientras que el resto se considera conjunto de entrenamiento.

La media de acierto en la predicción de los conjuntos de validación es el acierto de la validación cruzada.

Nuestra implementación es la siguiente: el usuario proporciona un intervalo de valores para C y otro para γ con el espacio de cuadrículas. Entonces, todos los puntos de la cuadrícula (C, γ) se combinan para averiguar que pareja nos proporciona una mayor precisión de validación cruzada. El usuario utiliza el mejor parámetro para entrenar el conjunto completo de entrenamiento y generar el modelo final.

Para una implementación sencilla, consideramos cada SVM con parámetros (C, γ) como un problema independiente. Como son tareas diferentes, podemos resolverlos en paralelo.

Una vez encontrados los valores de C y γ , y habiendo generado el modelo final, lo evaluamos usando la herramienta `svmpredict`, que nos devuelve un conjunto de etiquetas de predicción y unos valores de acierto (Accuracy) que nos indican lo buena ó mala que ha sido nuestra implementación.

Parte III

Problema Médico

7. Descripción

En el estudio médico [5] se analizaba la capacidad de discriminación de una nueva unidad de medida conocida como First Post Pacing Interval ó FPPI, para diferenciar dos tipos de taquicardias: ventriculares (TV) y supraventriculares (TSV). En dicho estudio, compuesto de múltiples fases, se implantó un dispositivo cardiodesfibrilador sobre 50 pacientes, de diferentes edades, todos ellos con defecto estructural coronario, además que 8 de ellos estaban tomando drogas antiarrítmicas.

El dispositivo lanzaba una terapia conocida como ATP (Anti Tachycardia Pacing), que consiste en una serie de pulsos secuenciales de corrección antitaquicárdica: de 5 a 10 pulsos con una duración de 91 % del ciclo de longitud de la taquicardia. El ATP estaba programado para cualquier ritmo ≥ 240 ms.

A continuación definiremos en que consiste el FPPI:

En el momento en el que el corazón de uno de los pacientes comienza a trabajar con un ritmo demasiado alto, superior al umbral de una taquicardia sinusal, el CDI lanza el ATP para intentar controlar el ritmo cardiaco. Una vez lanzada la terapia, se analiza el ritmo del primer intervalo, que es conocido como FPPI.

La hipótesis en la que se basaron para realizar ese estudio consistía en lo siguiente:

Como explicamos anteriormente, en el momento en el que se produce una taquicardia, se forma un circuito cardiaco, que es penetrado por el ATP lanzado por el CDI. Debido a que la distancia entre el vértice ventricular derecho y el circuito taquicárdico ventricular es menor que la distancia al atrio, la corrección de un circuito ventricular puede conseguirse

con menos estímulos que cualquier arritmia supraventricular, por tanto, el FPPI permanece estable cuando se comparan las descargas más cortas y más largas únicamente en el caso de una TV.

Basándonos en el hecho de que la diferencia y el cociente entre el FPPI y la LCT debe ser menor en las TV que en las TSV, se estudió la manera de aplicarlo en la clasificación de las taquicardias, evitando usar un algoritmo de discriminación ya existente; puesto que además de requerir más cálculo, más retardo y consumir más batería del dispositivo, fallaban en un intervalo del 12 % al 27 % a la hora de clasificar las TSV.

En este estudio previo se consiguieron implementar 2 métodos diferentes para la clasificación de Taquicardias en los pacientes con CDI:

1. El primer método está basado en un análisis simple de la FPPI-V entre ráfagas consecutivas de 5 a 10 pulsos. Consiguió clasificar correctamente la totalidad de los episodios de TV y el 95 % de las TSV, una vez se aplicó la totalidad de la secuencia ATP. Es un muy buen resultado, pero tiene un factor en contra importante: la duración de la secuencia ATP completa. De todos modos, una comparación de 2 ráfagas de 5 y 10 pulsos clasificará correctamente el 100 % de las TV y el 90 % de las TSV. Con esos ajustes, el tiempo estimado para completar el lanzamiento del ATP tras la detección de la TV es menos de 10 segundos. Este método tiene una limitación que no es considerada importante debido a la escasa probabilidad de que suceda: no es aplicable a las TV de ciclo rápido (> 320 ms).
2. El segundo método es un algoritmo que clasifica correctamente los episodios espontáneos usando únicamente las variables correspondientes a 2 ráfagas consecutivas de 5 a 6 pulsos. La ventaja de este algoritmo con respecto al método anterior es que puede ser aplicado incluso a TV de ciclo rápido.

8. Resultados del estudio

Para el análisis de la FPPI-V y del FPPI, se usaron los datos que se obtuvieron de los pacientes en el momento de la implantación del CDI y los episodios espontáneos que se produjeron durante los 12 meses siguientes al alta hospitalaria. Detallamos a continuación dichos datos y los resultados obtenidos de su estudio.

TV inducidas en los pacientes. Se indujeron 96 episodios de TV, de los cuales, 64 fueron terminados por los pulsos (66 %). Se pudo observar una aceleración de TV en 5 de esos episodios, y en 2 de ellos no se pudo medir la FPPI-V debido a una carencia d captura ventricular durante esos pulsos. Todos los episodios no terminados por la terapia ATP fueron tratados con descargas de alta energía de manera exitosa.

TV espontáneas detectadas durante el seguimiento. Durante el seguimiento se registraron 144 episodios de TV en 19 pacientes. Se consiguieron terminar 81 de esos episodios con la terapia ATP (60 %), la mayoría de ellos por la descarga de 5 pulsos (50 episodios). Únicamente 2 episodios fueron acelerados debido a la terapia.(< 1.5 %).

El FPPI no pudo medirse en 8 de los episodios, debido a que no hubo captura local en un ATP descartado en la LC programada. Al igual que en las TV inducidas, aquellos episodios que no fueron terminados por la terapia de pulsos fueron tratados con éxito por descargas de alta energía.

TSV espontáneas detectadas durante el seguimiento. Durante el seguimiento, se registraron 54 episodios de TSV en 12 pacientes, incluyendo 4 episodios de fibrilación atrial. De estos episodios, un 46 % fueron ralentizados gracias a la terapia de pulsos, y un 22 % fueron terminados, el 32 % restante no sufrieron cambio alguno. El ATP terminó los episodios que tenían una LC más corta que los episodios que fueron ralentizados.

Además, se dieron 11 episodios en 4 pacientes que recibieron descargas inapropiadas de alta energía. *La mayor parte de dichas descargas se produjeron durante taquicardias sinusales,*

taquicardia atriales y fibrilaciones atriales, debido a la dificultad existente para diferenciarlas respecto a las TV por falta de un electrodo atrial.

8.1. Análisis de discriminación

8.1.1. Capacidad de discriminación de la FPPI-V.

La FPPI-V era más baja en las TV que en las TSV expresadas en función de variabilidad absoluta, al igual que lo era el porcentaje de episodios en los que la FPPI-V era ≤ 50 ms. entre ráfagas consecutivas de 5 a 10 pulsos. En la figura mostrada a continuación se puede observar como los porcentajes de episodios con una variabilidad ≤ 50 ms en $T \geq 320$ ms: 100 % de las TV pero sólo un 5 % de las TSV tuvieron una FPPI-V ≤ 50 ms. De toda la comparación entre ráfagas de 5 a 10 pulsos, el FPPI se mantuvo estable tras todas las ráfagas ATP en las TV, pero cambió de manera continua en las TSV; por tanto, la FPPI-V era más corta en las TV cuando comparamos sólo 2 ráfagas de 5 y 10 pulsos.

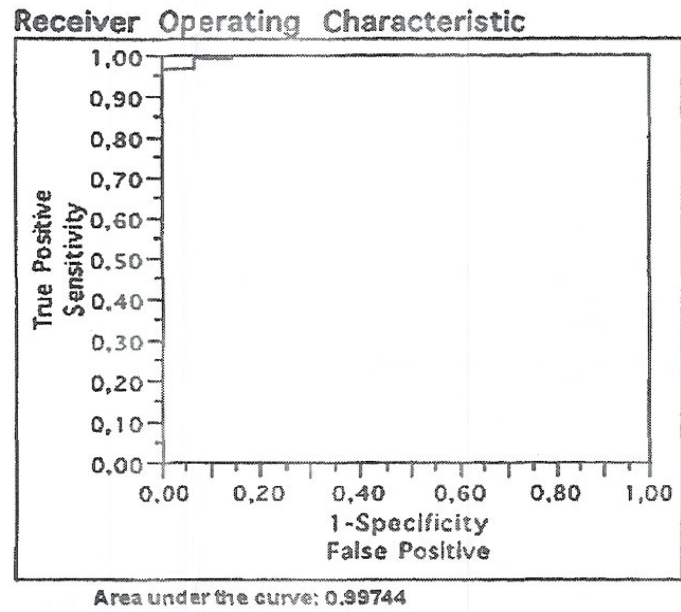
8.1.2. Capacidad de discriminación de parámetros relacionados con el FPPI.

Tanto el resultado de FPPI-TCL como el cociente de FPPI/TCL eran significativamente más bajos en las TV que en las TSV. El primer paso del análisis discriminatorio de TV / TSV se realizó en 176 episodios no finalizados por ráfagas de 5 ó 6 pulsos: 144 fueron TV y 43 TSV. La regresión logística multivariante inversa identificó a $FPPI(5)/TCL$, $FPPI(6)/TCL$ y a $FPPI(6)-TCL$ como medios independientes para predecir la discriminación de taquicardias. El añadido de estas variables permitió la creación de la ecuación predictora siguiente:

Algorithm 1 Algoritmo predictivo para la detección de Taquicardias Ventriculares.

$$P(TV) = e^{-35,11+16,35*FPPI(5)/TCL-67,75*FPPI(6)/TCL+0,22*FPPI(6)-TCL}$$

Sólo 2 de las TV inducidas que fueron terminadas tras una ráfaga más larga de 6 pulsos fueron clasificadas equivocadamente como TSV. Todos los episodios espontáneos se clasificaron correctamente.



$$P(VT)=e^{-35.11+16.35*FPPI(5)/TCL-67.75*FPPI(6)/TCL+0.22*FPPI(6)-TCL}$$

Figure 3 Predictive equation and Receiver Operating Characteristic (ROC) curve of the predictive model developed using FPPI-related parameters after 5 and 6 pulses.

Figura 15: Gráfica representando la probabilidad de detección y de falsa alarma del algoritmo predictivo

Imagen obtenida de [5]

Se puede observar que, usando únicamente las variables correspondientes a las descargas de 5 y 6 pulsos, obtenemos un clasificador con un porcentaje de acierto muy alto; la sensibilidad es máxima, mientras que la especificidad se mantiene siempre por encima de 90 %.

Parte IV

Plan de Pruebas

Durante esta parte del proyecto, realizamos una serie de pruebas iniciales para validar nuestro conjunto de datos, y al mismo tiempo, averiguar si la máquina SVM que utilizaríamos para evaluar nuestros resultados obtenidos a partir de los algoritmos de selección de características y de relación causal de variables era válida ó no.

Siguiendo esta idea, en primer lugar limpiamos la base de datos de muestras que no tuvieran todos los parámetros almacenados y tratamos las restantes normalizándolas.

En segundo lugar, replicamos el experimento previo con diferentes tipos de SVM intentando mantener el umbral fijado por la regresión básica usada en el estudio '**Differentiation of Ventricular and Supraventricular Tachycardia based on the analysis of the first postpacing interval after sequential anti-tachycardia pacing in implanted cardioverter-defibrillator patients**' tal como se muestra en la figura 9.

En tercer lugar, usamos la selección de características, obteniendo los Mantos de Markov con las características que cada algoritmo considera como relevantes y evaluamos dichos resultados realizando una clasificación de muestras.

Por último, aplicamos las relaciones de causalidad para obtener la estructura completa de los árboles, tanto de condiciones suficientes como del conjunto completo. Estos resultados no podemos validarlos, así que se entregarán al Doctor Arenal para su evaluación.

Detallamos cada punto de este esquema en mayor profundidad:

9. Filtrado de datos válidos

Durante la primera parte del proyecto tuvimos que hacer una limpieza de nuestra base de datos. Debido a la cantidad de información perdida en la recolección de parámetros para el

estudio, existían gran cantidad de observaciones que tuvimos que eliminar debido a que no tenían todos los valores transcritos. Se decidió eliminar cualquier observación que no tuviera todos los parámetros a utilizar en el experimento.

Antes de empezar a desarrollar la relación causal de variables, decidimos replicar el experimento realizado en [5] usando la misma base de datos, pero utilizando una SVM con diferentes tipos de clasificador en lugar de una regresión logística básica. Se realizaron una serie de pasos previos antes de cada tipo de clasificador que generalizamos poniéndolos a continuación:

1. Separamos nuestro conjunto de muestras, manteniendo la proporción de cada tipo, en dos subconjuntos diferentes : **entrenamiento** y **evaluación**.
2. Normalizamos las muestras de la siguiente manera: a cada muestra le restamos la media del conjunto de la variable, y dividimos todo por la varianza del mismo.
3. Asignamos un peso a cada clase en función de la cantidad de muestras respecto al total: a mayor número de muestras, menor peso de cada una:

$$\text{pesoC0} = \frac{\sum n_0}{\sum n_1} \quad \text{pesoC1} = \frac{\sum n_1}{\sum n_0}$$

10. Configuraciones

Los diferentes tipos de configuración se detallan a continuación:

Parámetros\Tipo de SVM	Lineal	Polinomio grado 1	Polinomio grado 2	Gaussiano
Coste C	C_L	C_{P1}	C_{P2}	C_G
Tipo de kernel: 't'	1	3	3	2
Gamma :'γ'	X	γ_1	γ_2	γ_G
Grupos de validación: 'v'	4	4	4	4
Ponderación Clase 0 'C_0'	pesoC0	pesoC0	pesoC0	pesoC0
Ponderación Clase 1 'C_1'	pesoC1	pesoC1	pesoC1	pesoC1
Grado del kernel: 'd'	X	1	2	X
Coeficiente 0: 'Coef0'	X	-0.25	-0.5	X

Cuadro 1: Tabla de Parámetros configurados para las clasificaciones SVM en la evaluación de resultados.

X: Indica que no se da valor a este parámetro en este caso.

Nota: En todos los casos, nuestra SVM está configurada como clasificación, luego omitimos el parámetro -s 0.

Los diferentes clasificadores quedan con los siguientes vector de opciones:

Tipo de Clasificador	Vector de opciones para introducir en la SVM
Lineal	options = '-s 0 -t 0 -c C -w1 pesoC0 -w1 pesoC1'
Polinomio grado 1	options = '-s 0 -t 3 -d 1 -r -0.25 -c C -g γ -w1 pesoC0, -w1 pesoC1'
Polinomio grado 2	options = '-s0 -t 3 -d 2 -r -0.5 -c C -g γ -w1 pesoC0 -w1 pesoC1'
Gaussiano	options = '-s 0 -t 2 -c C -g γ -w1 pesoC0 -w1 pesoC1'

Cuadro 2: Tabla con los vectores de opciones para las clasificaciones.

Observamos que un clasificador lineal y un polinómico de grado 1 son realmente parecidos, pero existen dos diferencias relevantes:

1. El clasificador lineal no permite aplicar la anchura de las gaussianas mientras que el polinomial de grado 1 si.

2. El clasificador polinomial contiene un factor ρ_0 para desplazar la frontera que no consta en el lineal.

11. Aplicación de algoritmos de selección de características

En el siguiente paso aplicamos la selección de características para obtener que factores eran relevantes para la clasificación proporcionando nueva información y cuales eran redundantes y podían ser eliminados de nuestro clasificador para reducir la carga computacional.

En este apartado tuvimos que seleccionar que algoritmos de todos los candidatos eran mejores para nuestro problema. Basándonos en el artículo [1], y en la documentación que venía incluida en el software de causalidad, decidimos dividir nuestro trabajo en dos subgrupos: el primero para el subconjunto previo y el segundo para el conjunto total de características.

Puesto que las variables tenían un gran rango de valores, optamos por utilizar aquellos algoritmos que trabajaban con variables continuas. De todos ellos seleccionamos dos: El **IAMB** y el **HITON_MB**, con la perspectiva de encontrar diferentes resultados, tal como se explica en el artículo '**Causal feature Selection**'.

Antes de comenzar tuvimos que realizar un tratamiento a los datos que íbamos a emplear para nuestra selección de características :

1. **Filtrar** las muestras de la base de datos original eliminando todas aquellas que no tuvieran todas las variables evaluadas.
2. **Normalizamos** las muestras de la siguiente manera: A cada muestra le restamos la media del conjunto de la variable en cuestión, y dividimos el resultado por la varianza del conjunto completo.

Una vez hecho esto, configuramos cada algoritmo según los parámetros deseados, tal y como mostramos a continuación:

Parámetros/Tipo de algoritmo	IAMB	HITON_MB
Matriz de datos	'indep_2c_6variables.xls'	'indep_2c_6variables.xls'
Variable Objetivo	tipo_arritmia (1)	tipo_arritmia (1)
Rango de valores		
Test de independencia usado	'z'	'z'
Umbral de independencia	threshold (i)	threshold (i)
Número Límite de variables del MB	X	-1

Cuadro 3: Tabla de parámetros para los algoritmos de selección de características, conjunto previo.

Parámetros/Tipo de algoritmo	IAMB	HITON_MB
Matriz de datos	'FPPI_indep.xls'	'FPPI_indep.xls'
Variable Objetivo	tipo_arritmia (1)	tipo_arritmia (1)
Rango de valores		
Test de independencia usado	'z'	'z'
Umbral de independencia	threshold (i)	threshold (i)
Número Límite de variables del MB	X	-1

Cuadro 4: Tabla de parámetros para los algoritmos de selección de características, conjunto completo.

- Rango de valores de las variables: '||' (Vacío para indicar que son continuas).
- Tipo de test de independencia a utilizar: 'z' (Test de Fisher-Z para variables continuas).
- Umbral de independencia: 'threshold (i)' (Usamos un vector para pasar diferentes valores de umbral).
- Límite de variables que componen el Manto de Markov: '-1' (Para indicar que no hay límite).

Quedan configurados de la siguiente manera:

IAMB	mb_1=Causal_Explorer('IAMB', data_norm, 1, [], 'z', threshold (i));
HITON MB	mb_2=Causal_Explorer('HITON_MB', data_norm, 1, [], 'z', threshold (i), -1);

Cuadro 5: Expresiones para el cálculo de MB para el conjunto previo.

Nos devuelve un vector con las variables que componen el Manto de Markov de la variable objetivo (tipo_arritmia) del conjunto de condiciones suficientes.

IAMB	mb_1=Causal_Explorer('IAMB',data_norm,1,[],'z',threshold (i));
HITON_MB	mb_2=Causal_Explorer('HITON_MB',data_norm,1,[],'z',threshold (i),-1);

Cuadro 6: Expresiones para el cálculo de MB para el conjunto completo de variables.

Nos devuelve un vector con las variables que componen el Manto de Markov de la variable objetivo (tipo_arritmia) del conjunto completo de variables.

11.1. Relación causal de variables

Por último aplicamos la relación causal de variables para establecer relaciones causa-consecuencia entre los diferentes factores conocidos, con la finalidad de encontrar un árbol causal completo.

Para ambos algoritmos, necesitamos configurar 7 parámetros los cuales desglosaremos a continuación, indicando de donde procede:

1. Los datos usados para la matriz de entrenamiento: 'data_norm'. (Datos normalizados procedentes de indep_2c_6variables.xls).
2. Rango de valores en los que se definen los datos del apartado anterior: '[]' (Vacío para indicar que son continuos).
3. Estadístico usado para medir la independencia: 'z' (Test-Z de Fisher para variables continuas).

4. Umbral máximo de medida para el test de independencia hallado a través de validación cruzada.
5. Flag indicando si usamos un superconjunto condicionado ó no: '1' (Lo cumple).
6. Flag indicativo si los datos cumplen la condición de Fidelidad: '0' (Indica que no la cumplen).

Parámetros/Algoritmos	PC	TPDA
Datos	indep_2c_6variables.xls	indep_2c_6variables.xls
Rango	[]	[]
Estadístico	'z'	'z'
Umbral	0.05	0.05
Superconjunto condicionado	1	1
Fidelidad	0	0

Cuadro 7: Tabla de asignación de valores para la relación causal de variables aplicado al conjunto previo

Y los algoritmos para hallar los árboles causales quedarían de la siguiente forma:

PC	A = Causal_Explorer('PC', data_norm, [], 'z', 0.05, 1, 0);
TPDA	B = Causal_Explorer('TPDA', data_norm, [], 'z', 0.05, 1, 0);

Cuadro 8: Expresiones de los árboles causales para el conjunto previo.

A continuación mostramos en la tabla la configuración para hallar los árboles causales completos de todo el conjunto de variables.

Parámetros/Algoritmos	PC	TPDA
Datos	FPPI_indep.xls	FPPI_indep.xls
Rango	[]	[]
Estadístico	'z'	'z'
Umbral	0.05	0.05
Superconjunto condicionado	1	1
Fidelidad	0	0

Cuadro 9: Tabla de asignación de valores para la relación causal de variables aplicado al conjunto completo.

Y los algoritmos para hallar los árboles causales completos quedarían de la siguiente forma:

PC	A = Causal_Explorer('PC', data_norm, [], 'z', 0.05, 1, 0);
TPDA	B = Causal_Explorer('TPDA', data_norm, [], 'z', 0.05, 1, 0);

Cuadro 10: Expresiones de los árboles causales para el conjunto completo.

Los resultados completos se entregarán a expertos médicos para ver si tenían sentido, lógica e interés desde el punto de vista médico para una futura línea de investigación.

Parte V

Resultados experimentales del proyecto.

12. Réplica del experimento de clasificación

12.1. Réplica usando SVM

En la réplica del experimento que hicimos en el primer paso del proyecto, utilizamos como base una SVM (Support Vector Machine), aplicándola como clasificador de varios tipos: gaussiano, polinomial, y lineal, para ver qué modelo obtenía mejores resultados, y, al mismo tiempo, utilizaba menor consumo recursos haciendo menos cálculos más rápidamente, con lo que se aplicaría la terapia de corrección de taquicardias mucho antes previniendo daños al paciente:

1. En primer lugar aplicamos un clasificador de kernel gaussiano sobre las cinco variables identificadas por el estudio [5] como conjunto previo para identificar el tipo de Taquicardia de manera sencilla y eficaz; dichas características son las siguientes: **FPPI 1 (5), FPPI 1 (6), FPPI 2 (6), TCL (5) y TCL (6)**.

12.1.1. Condiciones suficientes

Probabilidad de acierto	Sensibilidad	Especificidad
96.67 %	100 %	86.67 %

Cuadro 11: Resultados obtenidos con el conjunto previo.

2. A continuación aplicamos el mismo clasificador a todas las variables existentes (en este caso, 23), y vimos la mejoría que se producía respecto al caso anterior.

12.1.2. Todas las variables disponibles

Probabilidad de acierto	Sensibilidad	Especificidad
100 %	100 %	100 %

Cuadro 12: Resultados obtenidos a través del conjunto completo de variables.

Observamos que la SVM es sólida frente a problemas con muchas variables. Del mismo modo, comprobamos que existe información en las variables que no forman parte del conjunto previo, que sirve para afinar la especificidad del clasificador, permitiendo aumentar ésta última y la probabilidad de acierto total.

12.2. Evaluación del conjunto de condiciones suficientes

En este apartado, aplicamos diferentes tipos de clasificador para el conjunto de condiciones suficientes, obteniendo los siguientes resultados:

Clasificador	Probabilidad de acierto	Sensibilidad	Especificidad
Lineal	93.33 %	100 %	73.33 %
Gaussiano	98.33 %	100 %	92.31 %
Polinomial grado 1	96.67 %	100 %	85.71 %
Polinomial grado 2	98.76 %	100 %	96.67 %

Cuadro 13: Resultados experimentales según el tipo de clasificador el conjunto previo.

12.3. Evaluación del conjunto completo de variables

En este apartado, aplicamos diferentes tipos de clasificador para el conjunto completo de variables, obteniendo los siguientes resultados:

Clasificador	Probabilidad de acierto	Sensibilidad	Especificidad
Lineal	97.44 %	100 %	87.50 %
Gaussiano	100 %	100 %	100 %
Polinomial grado 1	97.44 %	100 %	87.50 %
Polinomial grado 2	100 %	100 %	100 %

Cuadro 14: Resultados experimentales según el tipo de clasificador el conjunto completo.

Debemos resaltar una serie de factores que condicionan la interpretación de estos resultados:

- El conjunto de muestras esta distribuido principalmente en dos conjuntos, permitiendo al clasificador lineal obtener muy buenos resultados, aunque no llegan a ser linealmente separables.
- La comparación entre nuestros resultados y los obtenidos por una regresión típica difieren únicamente en pequeños porcentajes en la especificidad, ya que nuestro clasificador consigue más precisión a cambio de una mayor complejidad de cálculo.
- De acuerdo con el estudio [5] es necesario mantener la sensibilidad al máximo aunque ello implique clasificar erróneamente algunas TSV como TV, debido a la gravedad de las últimas.

Algoritmo IAMB	Algoritmo HITON_MB
FPPI 1 (5), FPPI 1 (6)	FPPI 1 (5), FPPI 1 (6), FPPI 2 (6) PCL (5) y PCL (6)
Probabilidad acierto = 91.67 %	Probabilidad acierto = 98.76 %

Figura 16: Resumen de la selección de características.

13. Resultados experimentales de la selección de características

Antes de presentar cada experimento por separado, presentamos una tabla resumiendo los resultados de la selección de características aplicado al conjunto previo para cada método.

13.1. Aplicando el algoritmo IAMB al conjunto previo

Con el algoritmo IAMB obtuvimos que, del conjunto de las “condiciones suficientes”, sólo eran necesarias 2 de ellas: **FPPI 1 (5)** y **FPPI 1 (6)**. Aplicando el clasificador gaussiano con estas variables, los resultados que obtuvimos fueron los siguientes:

Clasificador	Probabilidad de acierto	Sensibilidad	Especificidad
Lineal	86.67 %	100 %	42.86 %
Gaussiano	91.67 %	100 %	64.29 %
Polinomial grado 1	88.33 %	97.78 %	53.33 %
Polinomial grado 2	90.00 %	100 %	57.14 %

Cuadro 16: Resultados de la validación de la selección de características IAMB.

13.2. Aplicando el algoritmo HITON_MB al conjunto previo

Con el algoritmo HITON_MB, se identificaron todas las variables como características relevantes (**FPPI 1 (5)**, **FPPI 1 (6)**, **FPPI 2 (6)**, **PCL (5)** y **PCL (6)**) obteniendo los siguientes resultados:

Clasificador	Probabilidad de acierto	Sensibilidad	Especificidad
Lineal	93.33 %	100 %	73.33 %
Gaussiano	98.33 %	100 %	92.31 %
Polinomial grado 1	96.67 %	100 %	85.71 %
Polinomial grado 2	98.76 %	100 %	96.67 %

Cuadro 17: Resultados de la validación de la selección de características HITON_MB

13.3. Aplicando el algoritmo IAMB al conjunto de completo de variables

Obtuvimos un resultado extraño: La única variable que consideraba era el FPPI (9), dando lugar a unos resultados de muy baja calidad, con un alto porcentaje de error.

Consultando diversos estudios y documentación [1], [2], [3], llegamos a la conclusión de que el algoritmo IAMB no funciona correctamente al tener un gran número de variables con tan pocas observaciones; de ahí que se tenga en cuenta el empleo del algoritmo HITON_MB, en estos casos; tal como se enuncia en el artículo '**Causal Feature Selection**'.

13.4. Aplicando el algoritmo HITON_MB al conjunto completo de variables

Se identificaron todas las variables como componentes del Manto de Markov, cumpliendo con el estudio [5]. Los resultados que obtuvimos fueron los siguientes:

Clasificador	Probabilidad de acierto	Sensibilidad	Especificidad
Lineal	97.44 %	100 %	87.50 %
Gaussiano	100 %	100 %	100 %
Polinomial grado 1	97.44 %	100 %	87.50 %
Polinomial grado 2	100 %	100 %	100 %

Cuadro 18: Resultados de la validación de la selección de características HITON_MB al conjunto completo

14. Resultados de la obtención del árbol causal

Para la obtención de los árboles causales usamos el software encontrado en Causal Explorer [2] y conseguimos la matriz de Adyacencia, y utilizamos el software Graphviz [11] para representarlos gráficamente.

NOTA: En el experimento previo [5] se usaron como conjunto previo las características **FPPI 1 (5)**, **FPPI 1 (6)**, **FPPI 2 (6)**, **TCL (5)** y **TCL (6)**, mientras que en nuestros resultados obtuvimos que las características eran **FPPI 1 (5)**, **FPPI 1 (6)**, **FPPI 2 (6)**, **PCL (5)** y **PCL (6)**. Hicimos una comprobación de correlación entre PCL y TCL obteniendo un valor de 0.999.

Ésto nos indica que la información que obtenemos de ambas variables es equivalente, por lo que su sustitución es totalmente comprensible.

Aplicamos dos algoritmos diferentes a cada conjunto de características: al conjunto previo y al conjunto completo de variables. Obtuvimos los siguientes resultados y figuras:

14.1. Algoritmo TPDA aplicado al conjunto completo de variables

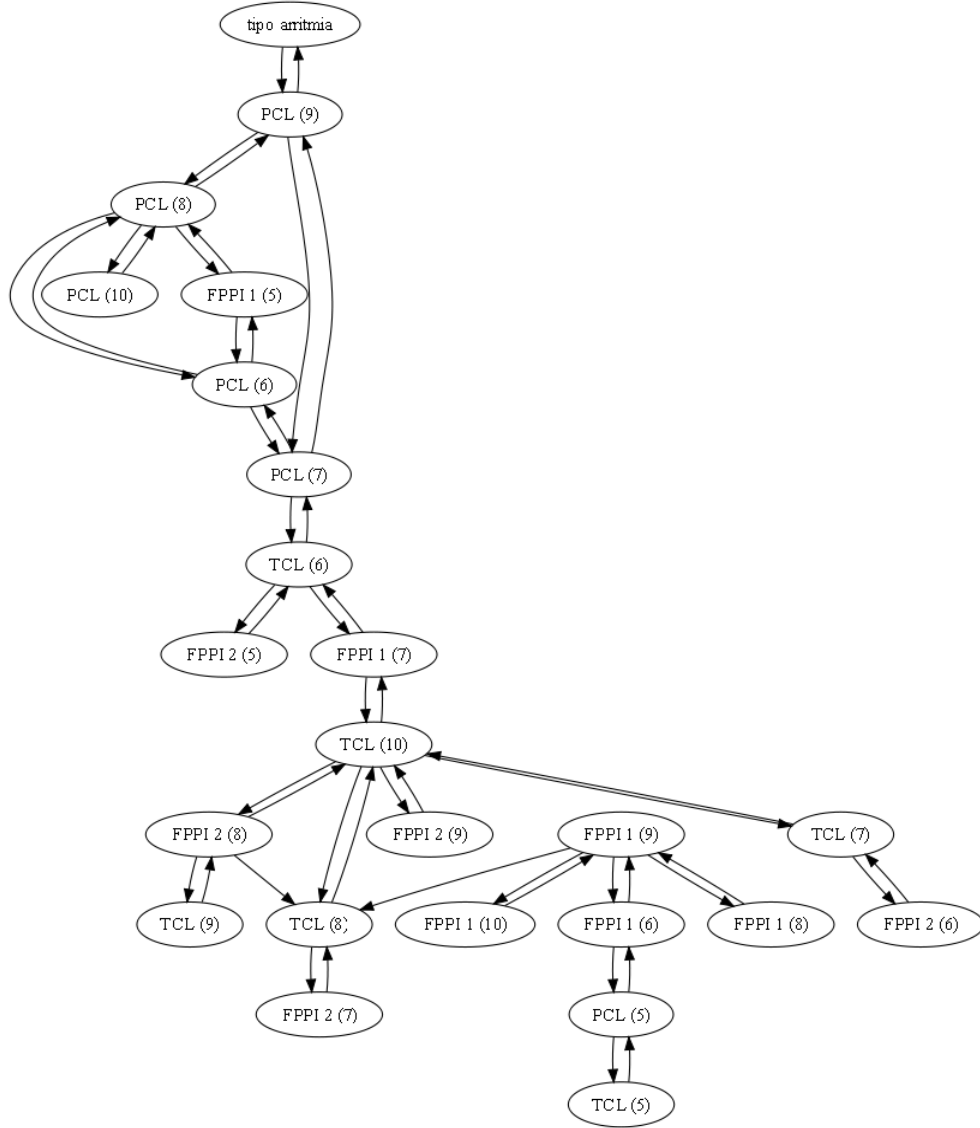


Figura 17: Árbol causal para el conjunto completo de variables usando el algoritmo TPDA.

En este caso, observamos que tipo arritmia tiene relación directa con PCL (9) a pesar de que no pueda orientarse. PCL (9) está relacionada con PCL (8) que suponemos será consecuencia de éste, al igual que con PCL (7).

Vemos relaciones que tienen sentido, por ejemplo FPPI 2 (9) y FPPI 2 (8) con TCL (10) ó FPPI 1 (8) con FPPI 1 (9), y este a su vez con FPPI 1 (10). A pesar de que no estén dirigidas, orientarlas de esta manera es usar la relación temporal. Se produce antes FPPI (9) que el FPPI (10), luego la orientación debe ser del primero al segundo.

Destacar que la gran mayoría de las relaciones están sin orientar debido a la escasez de muestras.

En este caso también observamos relaciones erróneas FPPI 1 (9) como causa de TCL (8) no tiene ningún sentido

14.2. Algoritmo PC aplicado al conjunto completo de variables.

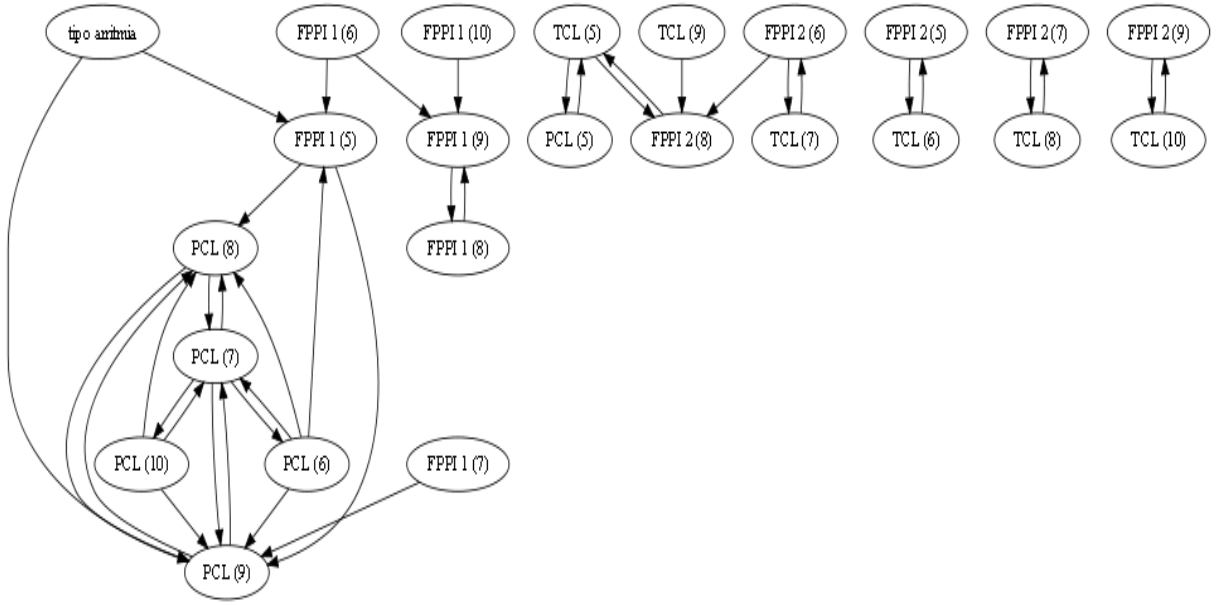


Figura 18: Árbol causal para el conjunto completo de variables usando el algoritmo PC.

En el caso del árbol completo observamos como tipo arritmia causa FPPI (6) y PCL (9), que tiene sentido ya que desencadena el ATP. Al lanzar la secuencia completa, la medida PCL es causada por el tipo arritmia. Sin embargo, observamos que FPPI 1 (6) es causa de FPPI

1 (5) lo cual no tiene sentido ya que es posterior a dicha medida. Del mismo modo no tiene sentido que FPPI 1 (10) sea causa de FPPI 1 (9) ó que TCL (9) sea causa de FPPI 2 (8).

Vemos también que gran parte de las unidades TCL están descolgadas del árbol. Nos indica que necesitamos más muestras para poder establecer el árbol completo, ya que no tiene sentido que en el apartado de las 6 variables formen parte del árbol llegando a ser causas directas, y que en este caso no formen parte del conjunto.

14.3. Algoritmo TPDA aplicado al conjunto previo

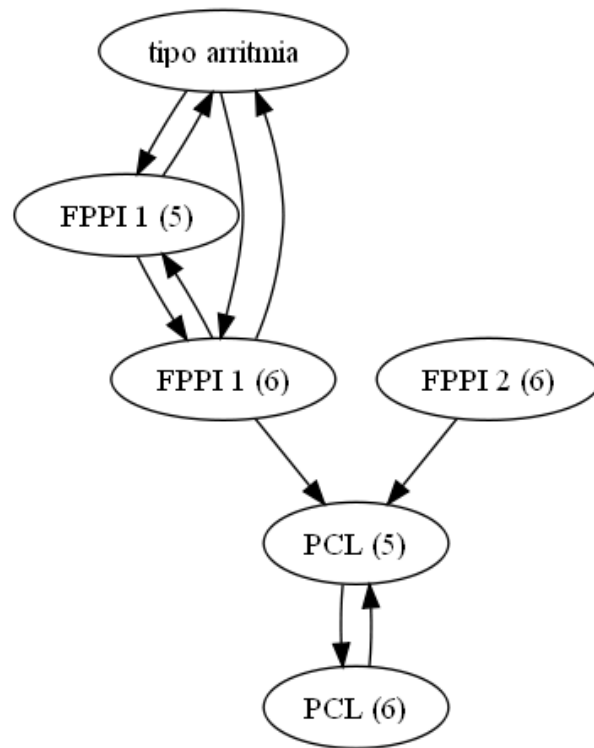


Figura 19: Árbol causal para las condiciones suficientes usando el algoritmo TPDA.

Podemos observar que en este caso tipo arritmia está relacionada de manera directa con FPPI 1 (5) y FPPI 1 (6). Lo lógico sería que fuera causa de ambas por el mismo motivo explicado arriba, que son medidas tomadas tras la aplicación del ATP.

Observamos que FPPI 2 (6) es esposa de FPPI 1 (6) respecto a PCL (5), lo cual a primera vista no resulta demasiado claro, considerando además que existe una relación no definida con PCL (6)

14.4. Algoritmo PC aplicado al conjunto previo

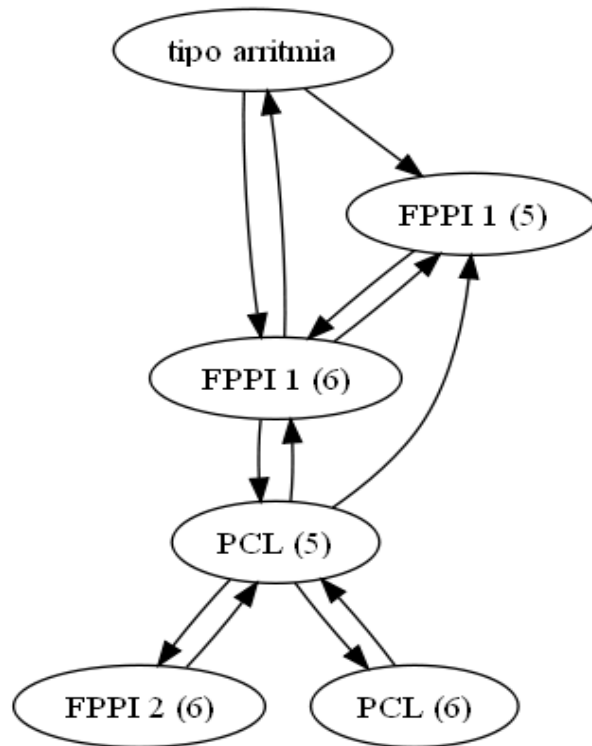


Figura 20: Árbol causal para las condiciones suficientes usando el algoritmo PC.

Observamos que en este árbol, tipo arritmia es causa de FPPI (5) y posiblemente de FPPI (6), lo cual tiene lógica teniendo en cuenta que las 2 medidas se producen al aplicar la terapia ATP en el momento de detectar la taquicardia.

Del mismo modo vemos que PCL (5) es causa de FPPI (5). Al ser una unidad de medida de ciclo cardiaco tiene sentido que sea causa, aunque también podría ser consecuencia ó relación bilateral.

Por último las relaciones de PCL (5) con FPPI 2 (6) y PCL (6) no han podido establecerse de manera definitiva. En este caso, necesitamos esperar a una interpretación médica para saber si tienen sentido.

Parte VI

Conclusiones

Después de todas las pruebas realizadas, aplicando distintas combinaciones de herramientas y obtener los resultados mostrados en el apartado anterior, llegamos a las siguientes conclusiones:

1. La aplicación de una SVM para replicar el experimento previo ha sido una elección acertada por los siguientes motivos:
 - a)* Conseguimos mantener como resultado mínimo los obtenidos por el estudio previo a la hora de clasificar las diferentes clases de taquicardias, pudiendo manipular más parámetros, ajustándonos a la situación real. Entre otros factores, pudimos ponderar el coste en función del tipo de error cometido y, al analizar cada muestra como una distribución gaussiana, tuvimos en cuenta la posible desviación del valor debido a errores de medida de los dispositivos.
 - b)* Así mismo, en media, obtuvimos unos resultados más precisos a la hora de clasificar las Taquicardias Supraventriculares, que era donde la precisión de la regresión básica no era tan acertada. A pesar de que la especificidad no es tan decisiva como la sensibilidad hemos podido aumentarla, llegando en algunos casos a alcanzar el 100 %.
 - c)* La carga computacional necesaria para mejorar dichas prestaciones en algunos casos es menor incluso que la realizada por la regresión. Al utilizar un kernel lineal, la transformación que aplicamos a las muestras es mucho más sencilla que la usada en el estudio [5]
2. A partir de los resultados obtenidos en la selección de características vemos que:
 - a)* Aplicando el algoritmo HITON_MB al conjunto de condiciones suficientes obtuvimos unos resultados similares a los enunciados en el estudio médico sobre la diferenciación de taquicardias, indicando que todas las características son relevantes

para la evaluación y clasificación de las mismas. A pesar de no poder extraer unas conclusiones determinantes debido a que necesitamos un punto de vista médico, el haber alcanzado este resultado nos indica que los Mantos de Markov obtienen resultados lógicos e interesantes.

- b)* Comprobamos así mismo la diferencia de aplicar el algoritmo IAMB con el HITON_MB; tal como se especifica en la documentación [1], el algoritmo HITON_MB proporciona mejores resultados al tener en cuenta relaciones no directas, dando lugar a Mantos de Markov más completos.

3. De los resultados de causalidad extraemos las siguientes conclusiones:

- a)* El tipo de arritmia condiciona el valor de los FPPI, manteniéndose más bajos en caso de que haya sido una taquicardia ventricular, tal como se demuestra en el estudio médico [5].
- b)* Existe una relación no definida entre los TCL tras las descargas de 5 y 6 pulsos de la terapia antitaquicárdica que se corresponde con la inserción de dichas variables en el Manto de Markov del apartado de selección de características. Dicha relación no tiene porque ser directa ni fuertemente relevante, pero sirve para ajustar mejor nuestra frontera permitiendo alcanzar mayores niveles de especificidad.
- c)* En determinadas relaciones causales, los algoritmos son incapaces de orientar los ejes, indicando que hay una relación a primera vista, pero que no podemos terminar de aclarar si es directa ó que podríamos eliminarla si ambas características compartieran antecesores comunes. Esta situación viene explicada con más detalle en el artículo [1].

Estas conclusiones son las que hemos podido extraer a partir de los resultados obtenidos desde el punto de vista de la ingeniería. Gracias a que podemos compararlo con lo que se obtuvo en el estudio anterior podemos evaluarlos como correctos, aproximados ó erróneos. Sin embargo, debemos tener en cuenta lo siguiente:

Tanto los algoritmos de selección de características como los de búsqueda de relaciones causales están configurados para trabajar con un alto número de variables y cientos de miles

de observaciones. En nuestro proyecto tuvimos que filtrar la base de datos, eliminando gran cantidad de observaciones debido a que no estaban registrados todos los parámetros, por lo que la muestra resultaba inútil.

Por tanto, a pesar de que concluimos de manera inicial que tanto la selección de características como la relación causal pueden emplearse en este tipo de situaciones, necesitaríamos más datos para poder afirmarlo de manera más rotunda.

Por último, faltaría una evaluación médica de los resultados para poder extraer conclusiones desde un punto de vista más fisiológico y ver posibles utilidades futuras y líneas de investigación.

Parte VII

Líneas futuras de investigación

En primer lugar nos interesaría modificar el grupo de trabajo, ampliándolo a un rango de mayores edades, incluyendo a pacientes que no tuvieran necesidad de tomar drogas antiarrítmicas ó que no tuvieran defecto congénito cardiaco.

Del mismo modo nos interesaría añadir alguna variable nueva de medida, como la distancia al punto de origen del circuito taquicárdico y ver su comportamiento una vez es aplicada la terapia ATP, así como nuevos tipos de información: edad de los pacientes, tipo de defecto congénito, efecto de las drogas antiarrítmicas sobre nuestras medidas, etc...

Una vez obtenidos estos resultados, y evaluados por los médicos, sería muy interesante observar el comportamiento de un CDI utilizando como clasificador de taquicardias una máquina SVM con kernel gaussiano que posee la probabilidad de acierto más alta con un número de variables mayor que el de condiciones suficientes, pero menor que el del conjunto completo.

Observando entre otras cosas el número de descargas inadecuadas, el gasto de batería del dispositivo, la velocidad de cálculo y el retardo de descarga, podríamos buscar la manera de optimizar los cardio-desfibriladores de manera estructural para mejorar aún mas las prestaciones.

Si se confirman los buenos resultados de la relación causal respecto a este estudio, se puede aplicar esta herramienta para hallar que factores contribuyen a la aparición de enfermedades degenerativas tales como Alzheimer, esclerosis y cáncer; hallando cuales son las causas directas para trabajar sobre ellas.

Aplicar los resultados obtenidos, tras una revisión por parte de los médicos del estudio [5], para estudiar las variables obtenidas como causas, y tratar de manera más efectiva los episodios de taquicardias ventriculares.

Fuera del plano médico, encontramos muchas utilidades posibles para esta herramienta, pero de manera resumida podemos expresarlo así:

En el momento en el que tenemos la estructura completa de relaciones causa-consecuencia, sabemos que factores causan una determinada consecuencia, por lo que podemos trabajar sobre ellos para cambiar el valor del resultado. Si por el contrario actuáramos sobre las consecuencias que produce dicho factor, no obtendríamos ningún resultado de interés, ya que no estaríamos tratando los factores que lo causan y no modificaríamos sus valores.

Parte VIII

Bibliografía

1. **Causal Feature Selection** by Isabelle Guyon, Clopinet, California Constantin Aliferis, Vanderbilt University, Tennessee Andrée Elissee, IBM Zürich, Switzerland, March 2, 2007.
2. **Causal Explorer: A Causal Probabilistic Network Learning Toolkit for Biomedical Discovery** by C.F. Aliferis, I.Tsamardinos, A.R. Statnikov, L.E.Brown from Vanderbilt University. May 2003
3. **Causal Explorer: Causal Discovery Algorithms** by Constantin Aliferis Alexander Statnikov Ioannis Tsamardinos. Vanderbilt University. May 2003
4. **An Algorithm For Generation of Large Bayesian Networks** by A. R. Statnikov, I. Tsamardinos, C.F. Aliferis Department of Biomedical Informatics, Discovery Systems Laboratory, Vanderbilt University Technical Report DSL-03-01 May 28th 2003.
5. **Differentiation of Ventricular and Supraventricular Tachycardia based on the analysis of the first postpacing interval after sequential anti-tachycardia pacing in implanted cardioverter-defibrillator patients** by Angel Arenal M.D, Mercedes Ortiz, PhD, Rafael Peinado, MD, Jose.L Merino, MD.
6. **Pattern Classification** by Duda, Richard. O. Third Edition. Editorial A Wiley-Interscience.
7. **Pattern recognition:** Theodoridis, Sergios. Third Edition. Editorial Academic Press.
8. **Prometheus, Atlas de Anatomia** by Gilroy, Anne Marie/ MacPherson, Brian R./ Ross, Lawrence M. 11^a Edición. Editorial Médica Panamericana.
9. **Kiega Blogspot** Laboratoristas clínicos de México del CBT. <http://kiegam5.blogspot.com>

10. **LIBSVM – A Library for Support Vector Machines** by Chih-Chung Chang and Chih-Jen Lin. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
11. **Graphviz - Graph Visualization Software.** <http://www.graphviz.org/>
12. **Curso Marcapasos SECEX** by Dr Enrique Fernández Burgos. <http://www.secex.org/marca/index.>
13. **Medtronic:** Imágenes e información sobre Taquicardias Ventriculares. <http://www.medtronic.com/E>
14. **Clínica DAM:** Servicio de Medicina. <http://www.clinicadam.com/>
15. **Electrocardiografía:** Imágenes e información sobre Taquicardias Sinusales.
16. **Texas Heart Institute.** Centro de información cardiovascular. www.texasheartinstitute.org/
17. **Harrison Principios de Medicina Interna** by Anthony S. Fauci, Eugene Braunwald, Dennis L. Kasper. 17^a edición. Editorial McGraw Hill
18. **Probabilistic Reasoning in Intelligent Systems** by Morgan Kaufmann, Pearl, J. 1 Edition. Editorial Morgan Kaufmann
19. **The support vector machine under test** by David Meyer, Friedrich Leisch, and Kurt Hornik. <http://www.jstatsoft.org/v15/i09/paper>

Glosario:

SVM: Support Vector Machine (ó máquina de vectores soporte).

TS: taquicardia sinusal.

TSV: taquicardia supraventricular.

TV: taquicardia ventricular.

FV: fibrilación ventricular

AV: aurico-ventricular.

GAD: grafo acíclico dirigido.

CMC: condición causal de Markov.

CFC: condición de fidelidad de Markov.

PC: Parent-Children.

SVC: Support Vector Classification.

SVR: Support Vector Regression.

TPDA: Three- Phase Dependency Analysis.

FPPI: First Post Pacing Interval.

CDI: cardioDesfibriladores implantados.

ATP: anti Tachycardia Pacing.

LCT: longitud de ciclo de la taquicardia.

FPPI-V: First Post Pacing Interval variability.

LC: longitud de ciclo.

IAMB: Iterative Associative Markov Blanket.

HITON_MB: HITON Markov Blanket.

PCL: Peak Cycle Length.

